Physical function and cognitive decline in the older population: an investigation on prospective associations and measurement challenges

Physical function and cognitive decline in the older population: an investigation on prospective associations and measurement challenges

Isabella Rizzo
MPhil Psychology
Birkbeck, University of London

With this statement I hereby declare that the submitted MPhil thesis is my own work

Abstract

Physical fitness is thought to be associated with optimal levels of cognition in older adults, and is being increasingly investigated as a modifiable risk factor for dementia. Given the growing rates of cognitive impairment in the ageing population, establishing the association between physical fitness and dementia holds great clinical significance. The current work systematically reviewed prospective cohort studies looking at the association between baseline performance-based measures of physical function and the subsequent development of cognitive decline or dementia diagnosis. Findings from our review showed longitudinal associations between most of the physical function variables that we considered and cognitive outcomes, whereby greater levels of physical activity were associated with reduced risk of cognitive decline and dementia. The importance of these findings was discussed, highlighting the value of these prospective associations for policy making and healthcare systems. Some recommendations were made to improve methodological rigour in future longitudinal research, in light of the methodological heterogeneity and inconsistent reporting that we observed in the studies that were retrieved. Another issue which emerged from our search was the predominant use of self-reported measures to assess physical fitness. The current work reviewed the literature investigating the consistency between self-reported and performance-based measures of physical function, and conducted a secondary data analysis of cross-sectional data, with the aim to investigate the agreement between self-reported and performance-based measures of physical functioning. To this end, a secondary data analysis was conducted looking at the correlation between a self-reported measure of physical fitness and three objective measures of physical functioning. Further, our secondary data analysis included stepwise regression analyses investigating cognitive outcomes from models based on age, a self-reported measure of physical fitness and a measure of grip strength. Our correlational analysis showed modest rates of agreement between self-reported and performance-based measures of physical fitness, which were broadly in line with the correlations found by previous research. Results from the regression models showed that grip strength did not explain any additional variance in the cognitive outcome that was not already accounted for by the PFQ, suggesting that both measures are capturing the same construct. These findings were discussed in light of the concept of convergent validity, and some suggestions were made to address suboptimal convergence between self-reported and objective measures of physical fitness. Further, the use of regression analyses investigating the amount of overlapping variance explained by self-reported and performance-based measures was suggested as a method to validate self-reported questionnaires of physical fitness in future research.

Table of Contents

List of tables ... *6*

List of figures ... *6*

Chapter 1: Preventing dementia ... *8*

Introduction ... **8**

The ageing population ... **9**

Implications for healthcare ... **10**

Physiological changes in normal and pathological ageing ... **12**

Changes in the musculoskeletal system ... 12

Changes in the brain ... 13

Dementia and Alzheimer's disease: prevalence, aetiology and clinical presentation ... 15

Successful ageing and risk factors for dementia ... **16**

Physical fitness and cognitive trajectories ... 18

Establishing temporal associations: the significance of longitudinal evidence ... **20**

Performance-based measures of physical fitness in ageing research ... **21**

Chapter 2: Longitudinal association between performance-based physical fitness measures and cognitive decline: a systematic review of prospective studies ... *24*

Introduction ... **24**

Existing reviews ... **25**

Methods ... **27**

Search strategy ... 27

Eligibility criteria ... 27

Study selection and data collection ... 28

Flow diagram ... 29

Data extraction ... 30

Characteristics of the studies ... 35

Results ... **36**

Assessment of study quality and covariates ... 36

Meta-analysis ... 39

Narrative synthesis of the results ... 47

Discussion ... **56**

Chapter 3: Longitudinal association between performance-based physical fitness measures and dementia risk: a systematic review of prospective studies ... *65*

Introduction ... **65**

Methods ... **65**

Data extraction ... 67

Characteristics of the studies ... 69

Results ... **70**

Assessment of study quality and covariates ... 70

Meta-analysis ... 73

Synthesis of additional studies ... 77

Discussion ... **79**

Chapter 4: Performance-based and self-reported measures of physical fitness in ageing research *84*

Introduction **84**

Evaluating a complex construct **85**

Issues with the use of self-report measures in ageing research **86**

Secondary data analysis of objective and self-report physical fitness data of healthy individuals **90**

Consistency between objective and self-reported measures of physical fitness: existing evidence 90

Methodology of the studies 92

Data analysis 94

Results **96**

Discussion **109**

Chapter 5: Main findings and recommendations for future research *117*

Main findings **117**

Methodological and reporting inconsistencies in prospective cohort studies: promoting common guidelines in ageing research **118**

Designing longitudinal studies **119**

Longitudinal studies for the older population 119

Operationalisation of physical and cognitive outcomes 120

Analysis methods that capture the physical function-cognition association effectively 121

Establishing basic criteria for covariate adjustment 122

Defining follow-up duration 123

Open data sharing **123**

Mitigating the issue of reverse causality in longitudinal research **124**

Clinical implications of the association between baseline physical function and later cognitive decline/dementia incidence **127**

The issue of suboptimal convergent validity between self-reported and performance-based measures of physical function **131**

Future research **134**

Conclusion **136**

References *137*

Appendix A *152*

List of tables

Table 1: Data extraction of relevant study variables, for studies with a cognitive decline outcome *30*
Table 2: Assessment of study quality *37*
Table 3: Data extraction of relevant study variables, for studies with a dementia outcome *67*
Table 4: Assessment of study quality *71*
Table 5: Summary of the measures employed in the studies that were used for secondary data analysis *92*
Table 6: Descriptive statistics divided by study and for the overall sample. *96*
Table 7: Coefficients for the correlations between PFQ and grip strength divided by PFQ groups *99*
Table 8: Coefficients for the correlations between PFQ and grip strength divided by age groups *100*
Table 9: Regression coefficients for the stepwise regression with Age, PFQ and Grip onto MoCA. *105*
Table 10: Regression coefficients for the stepwise regression with Age, Grip and PFQ onto MoCA. *106*
Table 11: Regression coefficients for the stepwise regression with Age, PFQ strength and Grip onto MoCA. *106*
Table 12: Regression coefficients for the stepwise regression with Age, PFQ aerobic conditioning and Grip onto MoCA. *107*
Table 13: Regression coefficients for the stepwise regression with Age, PFQ balance and grip onto MoCA. *108*

List of figures

Figure 1: PRISMA flow diagram for the search of the systematic review *29*
Figure 2: Forest plot for the meta-analysis of walking speed and dichotomised cognitive decline *39*
Figure 3: Forest plot for the meta-analysis of walking speed and continuous cognitive decline *40*
Figure 4: Funnel plot for the analysis between walking speed and continuous MMSE decline *40*
Figure 5: Forest plot for the meta-analysis of walking speed and continuous cognitive decline *41*
Figure 6: Forest plot for the meta-analysis of handgrip strength and dichotomised cognitive decline *42*
Figure 7: Forest plot for the meta-analysis of handgrip strength and continuous cognitive decline *43*
Figure 8: Funnel plot for the analysis between handgrip strength and continuous cognitive decline *43*
Figure 9: Forest plot for the meta-analysis of accelerometer-assessed daily physical activity and AD8 .... *44*
Figure 10: Forest plot for the meta-analysis of balance and continuous cognitive decline *45*
Figure 11: Forest plot for the meta-analysis of chair-stands and continuous cognitive decline *46*
Figure 12: Forest plot for the association between handgrip strength and dementia risk *73*
Figure 13: Funnel plot for the association between handgrip strength and dementia risk *74*
Figure 14: Forest plot for the association between walking speed and dementia risk *74*
Figure 15: Forest plot for the association between walking speed and dementia risk *75*
Figure 16: Forest plot for the association between balance and dementia risk *75*
Figure 17: Forest plot for the association between timed up and go test and dementia risk *76*
Figure 18: Forest plot for the association between chair-stands test and dementia risk *76*
Figure 19: Scatterplot showing the correlation between age and handgrip strength *97*
Figure 20: Scatterplot showing the correlation between age and PFQ *97*
Figure 21: Scatterplot for the correlation between PFQ scores and handgrip strength *98*
Figure 22: Scatterplot showing the association between PFQ and handgrip strength in the higher PF group *99*

Figure 23: Scatterplot showing the association between PFQ and handgrip strength in the intermediate PF group..... 99

Figure 24: Scatterplot showing the association between PFQ and handgrip strength in the low PF group................ 100

Figure 25: Scatterplot showing the association between PFQ and handgrip strength in the younger age group......... 101

Figure 26: Scatterplot showing the correlation between a composite performance-based measure derived by averaging balance and handgrip strength measures, and composite PFQ scores.. 102

Figure 27: Scatterplot showing the correlation between a composite performance-based measure derived by averaging chair stands tests and handgrip strength measures, and composite PFQ scores... 103

Figure 28: Scatterplot showing the correlation between a composite performance-based measure derived by averaging balance, chair stands and handgrip strength tests measures, and composite PFQ scores 104

Chapter 1: Preventing dementia

Introduction

Global life expectancy has increased by six years from 2000 to 2019 (World Health Organisation (WHO), 2019), leading to an exponential growth of the adult population aged 65 and over (Murman, 2015). There are numerous healthcare and medical advancements, as well as socioeconomic factors underlying this remarkable trend. Namely, the improved effectiveness of healthcare and medical systems, and the increasing allocation of resources to education, housing and sanitation are believed to be significant determinants of wellbeing globally (Bloom & Canning, 2007). Together with other aspects of socioeconomic development such as improved access to social welfare, the above indicators are believed to be driving forces of increasing longevity. While an upward trend in life expectancy is an undoubtable index of societal progression, there are pressing issues to be addressed as the ageing population continues to grow, generating higher societal costs and soaring pressure on healthcare (Kuh et al., 2014). Indeed, the ageing population places strain on health and social care systems, as increased longevity is associated with the growing prevalence of chronic age-related diseases and multi-morbid conditions (Fried et al., 2001). In particular, the last two decades have seen a steep increase in dementia and dementia of the Alzheimer's type, the rates of which are in parallel with the world population growth rate (Alzheimer's Association, 2014). With figures showing that dementia prevalence is expected to triple by 2050 (Alzheimer's Association, 2014). and in absence of well-established treatment options available, the development of preventive strategies targeting modifiable risk factors for dementia to curb the onset or progression of cognitive decline in the elderly is acquiring increasing relevance. One such modifiable risk factor is physical fitness, an aspect of general lifestyle which has been widely demonstrated to hold a preventive power against cognitive decline, dementia onset and outcomes of this disorder.

This chapter will explore the difficulties faced by healthcare systems when meeting the increased demand caused by age-related diseases, as well as some of the societal and economic implications of the ageing population and the increasing rates of progressive neurocognitive disorders. An overview of the processes of physiological decline encountered in ageing will be outlined, with particular focus on the mechanisms that drive the brain deterioration underlying cognitive disorders, and explaining how these disorders may lead to functional dependence in the elderly. This chapter will then introduce emerging findings on risk factors for dementia, and subsequently focus on the application of modifiable risk factors for dementia to develop disease prevention strategies against cognitive deterioration in older adults. The current work will focus on the modifiable risk factor physical fitness, and discuss empirical evidence in support of its preventive potential against cognitive decline.

Physical fitness (PF) is defined as "the ability to execute daily life activities with optimal performance, endurance and strength" (Caspersen et al., 1985), and will be defined hereafter as a state of physical well-being that is positively contributed to by physical activity (PA), or any musculoskeletal movement that requires the expenditure of energy (Hollamby, Davelaar, & Cadar, 2017). In particular, this chapter will discuss the importance of employing longitudinal research designs to investigate the impact of risk factors on cognitive outcomes, in order to best capture the changing processes that characterise cognition. Finally, it will focus on challenges surrounding the use of subjective measures of physical activity in ageing research, highlighting the need to also adopt a more objective approach, employing performance-based measures to assess this construct.

The ageing population

The number of older people aged 65 years or over has seen a drastic increase in the last two decades, and is anticipated to double by 2050, when it is expected to reach 2.1 billion globally (WHO, 2019). By 2050, the older population are expected to outnumber younger people aged 10 to 24, with significant repercussions on the structure and functioning of modern societies (WHO, 2019). The phenomenon underlying this demographic shift is thought to be driven by a range of factors, which can be identified primarily as medical and societal determinants (Brown, 2015). The primary reason why people live longer is that they have easier access to more efficient health and medical care (Brown, 2015). Indeed, efficient healthcare systems granting timely access to treatment and diagnostic services have largely contributed to the achievement of better health outcomes and greater life expectancy in the elderly (Hao et al., 2020). Furthermore, advancements in medical research have allowed for age-related life-threatening conditions such as cardiovascular disease, cancer, and diabetes to be successfully treated, reducing mortality rates for these highly prevalent deadly diseases (Brown, 2015). Furthermore, improvements in healthcare and medical care have contributed to reduce child mortality, and have granted the access to immunisation programmes in early childhood: both these factors play a significant role in the greater longevity of the population (Rappuoli et al., 2014).

Societal advancements have also facilitated an increase in life expectancy: the globalisation of economies leading to higher average incomes and easier access to education are factors that have considerably contributed to this mutating demographic trend (Brown, 2015; Aburto et al., 2020). There are indeed well-established links between education and life expectancy, showing that highly educated individuals live nearly 20% longer than their less educated counterparts (WHO, 2015). Additionally, figures show that people with higher incomes have up to 37% greater life expectancy compared to individuals with lower paid jobs (WHO, 2015). This may be driven by the fact that prosperous people with higher education levels and greater financial possibilities may have prompter access to high-level medical care. Finally, advancements in infrastructures, such as improvements in sanitation and greater availability of affordable housing have enhanced general living conditions, leading to increased longevity (Brown, 2015). Other aspects underlying the notable demographic shift are the widespread availability of nutrient-dense foods and water supply, which have made it easier for people to conduct healthier lifestyles (Brown, 2015).

The medical and societal advancements discussed above can certainly be regarded as a remarkable index of progression for societies worldwide. However, as the demographics of the world population shift, there is a pressing urge for governments to adapt to the changing needs of the elderly population on a global scale. Indeed, this demographic change has considerable societal implications on how public services such as pensions, social and health care will be developed and distributed (WHO, 2019). Most importantly, the ageing population has an increased risk of ill-health and disability, which are more frequent at an old age. The increase in life expectancy has resulted in people dying in old age, frequently after experiencing chronic illness in old age. Chronic conditions, neurodegenerative disorders and multi-morbid states that are commonly associated with older age are thus highly prevalent, and will continue to grow in prevalence as the distribution of the population demographics changes. For instance, 30% of adults aged 60 or over in the UK receive a diagnosis of dementia before death (Brayne, 2007). Projections show that this percentage will continue to rise, as the prevalence of dementia is strongly correlated with older age (Alzheimer's Association, 2014). This is not only the case for dementia: the growing prevalence of multi-morbidities, chronic disorders and disability in older age represents a significant challenge for healthcare systems worldwide, and raises the question of whether the quality of life of the elderly has improved in parallel with their lifespan (Brown, 2015).

Implications for healthcare

The primary impact of the ageing population on healthcare systems is the greater utilisation of health services by the elderly (Brown, 2015). Older adults consume a significant portion of healthcare and hospitals, where they are estimated to occupy up to 50% of workload and capacity (Rechel et a., 2009). Accordingly, figures show that the percentage of healthcare expenditure dedicated to older adults amounts to approximately 30-50 % of overall health costs (Rechel et al., 2009). The healthcare consumption of older adults has considerable societal and economic repercussions: the cost of dementia for the UK's economy alone is currently estimated at £26 billion a year (Lewis et al., 2014), and over $817 billion globally (Alzheimer's Association., 2014), with projections indicating that these numbers are expected to continue to rise steeply as dementia prevalence keeps increasing.

Further, the higher prevalence of chronic conditions in older adults means that older patients will frequently need regular access to health care services such as physiotherapy, nursing, and personal care assistance. The increasing demand for such services means that health and social care systems will need to change the way that their supply is organised, in order to adapt to the constantly increasing demand (Rechel et al., 2009). Existing systems need to adapt the way that they operate, in order to facilitate the optimal management of chronic conditions, which frequently falls within the remit of long-term care facilities. In particular, the implementation of home-based care networks is of relevance, as it mitigates the overload of care facilities, and it helps meet the needs of functionally dependent individuals. Indeed, older adults affected by chronic conditions which prevents them from living autonomously may increasingly need to move to long-term care facilities, or to receive frequent domiciliary care (Alzheimer's society, 2014).

Of crucial importance in helping the existing systems meet the mutating healthcare demand will be the development and implementation of interventive strategies aiming at the prevention and management of chronic, disabling conditions (WHO, 2019). This strategy reduces the burden and the costs of the ageing population on healthcare, by promoting healthier lifestyles in older age. Early interventions may target obesity, hypertension, and cognitive deterioration, amongst other age-related conditions. The promotion of interventions targeting these conditions may facilitate healthy ageing and contribute towards reducing the effect of the ageing population on health care expenditure. One type of intervention that has gained increasing popularity in recent years and that will be a central topic of discussion in this chapter are programmes targeting physical fitness, which is thought to mitigate cognitive deterioration by preventing vascular damage to the brain (Harada et al., 2013). In order to understand the mechanisms by which this type of intervention may operate, the following section of this chapter will provide an overview of the main changes involved in physiological and pathological ageing, and how these may result in impaired cognitive and physical functioning. It will subsequently discuss how interventions targeting physical fitness levels may contribute towards preventing cognitive decline and dementia onset.

Physiological changes in normal and pathological ageing

Ageing is a normal aspect of life, resulting from the cumulative processes of cellular and molecular degeneration over time. These processes lead to gradual changes in physical and cognitive capacity, and may cause the emergence of debilitating chronic conditions or ill-health. Indeed, a variety of age-related diseases are caused by normal biological processes of physiological decline driven by age, which frequently result in progressive functional limitations for the elderly (Clouston et al., 2013). The process of physiological decline involves primarily the musculoskeletal, nervous and sensory systems, which may be differentially, and often unpredictably impacted by ageing. Here, we will examine the main age-driven physiological changes to muscle, bone and brain structure, which have significant implications for the daily life function of older adults, and are central aspects to the variables examined in the current work.

Changes in the musculoskeletal system

Normal ageing is denoted by a deterioration in physical capability, known as one's ability to perform physical tasks of daily living (Cooper et al., 2011). This deterioration is driven by spontaneous processes of bone mass reduction and muscle atrophy, and a simultaneous increase in adipose tissue (Keller et al., 2014). Physical strength is estimated to decline at the rate of approximately 3% per year after 60 years of age (Morley et al., 2001). The reduction in physical strength appears to be caused by a reduced ability of cells in muscle fibres to produce protein, leading to reduced muscle volume and power (Amarya et al., 2018). Aside from reduced strength, overall mobility is also compromised due to age-related stiffness and degeneration of the connective tissues inside joints, which are in charge of reducing friction between bones during movements (Amarya et al., 2018). Furthermore, bone mass also deteriorates as a result of ageing. Although this process is less understood, it is believed to be caused by imbalances in the process of bone remodelling, which replaces old bone with new bone: with ageing, the quantity of bone produced in each remodelling cycle decreases, eventually leading to reduced bone mass (Demontiero et al., 2012). Age-related conditions such as vascular or metabolic disorders, as well as vitamin deficiencies (mostly calcium and vitamin D) and hormonal influences may have an impact on bone mass loss (Padilla Colon et al., 2018). This is especially relevant for women, where menopause accelerates bone reabsorption due to reduced oestrogen levels. With reduced bone mass comes the increased risk of fractures, which has a significant impact for an age group already at high risk for falls.

The abovementioned qualitative and quantitative changes in bone and muscle structure cause greater physical frailty, increased risk of fractures, and gradual loss of functionality. These structural changes may indeed result in various functions involved in daily life activities such as walking or physical strength declining over time. The gradual loss of engagement in everyday life activities due to reduced physical capability may ultimately lead to functional dependency for the elderly, with significant implications for their quality of life (Clouston et al., 2013).

Changes in the brain

Ageing is associated with cognitive and neurological disorders, caused by the gradual loss of the ability of the brain to elaborate and transmit signals through efficient connections (Amarya et al., 2018). Disorders characterised by progressive cognitive deterioration, behavioural changes and functional decline such as dementia and Alzheimer's disease are becoming increasingly prevalent as life expectancy continues to grow. Ageing is characterised by changes in brain structure and function, which are the primary cause of the cognitive decline seen in these disorders. Amongst the structural changes underlying cognitive deterioration is widespread neural degeneration, resulting in a reduction in brain volume, which has been shown to decrease at a 5% rate of reduction for each decade after the age of 40 (Peters, 2006). The reduction in brain volume particularly affects the hippocampus, which is involved in the formation of new memories, and the frontal cortex, which regulates executive functions, allowing for humans to plan behaviour in a flexible manner (Harada et al., 2013). Further to brain volume changes, there is a decline in synaptic density and quality: the connections between brain cells are reduced or become less efficient, which results in the transmission of neural signals being compromised. Finally, an additional structural change at the root of cognitive dysfunction is the disruption of the white matter tracts connecting frontal areas of the brain, which results in less efficient cognitive processing abilities (Harada et al., 2013).

Changes in cognitive processing abilities are thought to be primarily mediated by a decline in the levels of neurotransmitters. Dopamine is a neurotransmitter that is involved in various behavioural and cognitive processes. Dopamine levels decline at a rate of 10% every ten years starting from early adulthood, meaning that the synapses in the dopaminergic pathways may become less efficient with age (Peters, 2006). Serotonin and brain-derived neurotropic factor (BDNF) are other neurotransmitters involved in neurogenesis (i.e. the formation of new neurons in the brain): the level of serotonin and BDNF declines with ageing, potentially explaining a reduction in the formation of new neural cells (Martinowich & Lu, 2008; Miranda et al., 2019). Acetylcholine is another neurotransmitter, which is greatly involved in brain cognitive function and memory (Vallianatou et al., 2019). Acetylcholine levels have been shown to decline with age, and the decrease in this neurotransmitter is related to the development of dementia and Alzheimer's disease (Vallianatou et al., 2019). The ageing brain is also subject to reduced cerebrovascular efficiency, mostly due to age-related small vessel disease, which leads to reduced blood flow in the brain (Peters, 2006). The reduced supply of blood flow to the brain causes suboptimal oxygen flow and glucose input to the brain, leading to damages of ischaemic nature, which may also partially explain cognitive changes in older adults (Peters, 2006).

The structural and functional changes discussed above are manifested as a decline in cognitive functioning, with different components of cognition such as attention, memory, processing speed and executive functioning gradually deteriorating from midlife onwards (Singh-Manoux, 2012). The progressive age-related decline in cognitive functioning is known as cognitive ageing and can cause the elderly to experience functional deterioration, as difficulties may be encountered remembering new information, planning activities and staying focused on tasks (Alzheimer's Society, 2014). These changes are often associated with risk of injury, decline in daily life functioning, and increased risk of death (Amarya et al., 2018). The predominant cognitive change associated with ageing is memory decline, which is usually the first complaint raised by older adults with cognitive decline, or by their relatives (Harada et al., 2013). This decline primarily involves declarative memory, which constitutes "the conscious recollection of facts and events" (Harada et al., 2013). Executive functions, which include functions such as planning, organising, reasoning and cognitive flexibility, are also significantly impacted by the process of healthy ageing (Harada et al., 2013). The significant extent of the decline observed in memory and executive function in older age may be due to the grey matter volume reduction described above being most prominent in the frontal cortex (Harada et al., 2013; Yuan & Raz, 2014) and in the hippocampus (Raz et al., 2005), which are significant neural substrates for executive functions and memory respectively. However, cognitive ageing may have heterogeneous profiles, and the decline may affect other areas of cognition, such as attention, processing speed, and visuospatial abilities. Cognitive ageing may also be manifested as Mild Cognitive Impairment (MCI), which defines an early stage of cognitive impairment that may affect memory or another cognitive ability such as spatial perception or language, where the individual is still able to maintain independent functionality (Harada et al., 2013).

Dementia and Alzheimer's disease: prevalence, aetiology and clinical presentation

As discussed so far, there is a certain degree of cognitive decline associated with normal ageing. Considering that widespread cognitive decline is quite frequent in older age, it may be difficult to discern the difference between cognitive deterioration and symptoms that may suggest the onset of dementia (Singh-Manoux & Kivimaki, 2010). As a general guiding principle, when cognitive changes become so severe that they compromise an individual's daily life, there are reasons to believe that these changes might be attributable to dementia (Alzheimer's Associaion, 2014). Dementia is a disorder of progressive nature, defined by the impairment of multiple cognitive functions, which is severe enough to significantly impact daily life functioning (Singh-Manoux & Kivimaki, 2010). Dementia may affect different cognitive functions such as memory, reasoning, language, judgement and orientation, and may also have an impact on behaviour, mental health and emotional control. Given the progressive nature of dementia, the cognitive deficits that characterise this disorder may become increasingly severe over time, leading to gradual functional dependence for those affected. The incidence rates of dementia and dementia of the Alzheimer's type are growing steeply in parallel with the rising longevity of the population (Alzheimer's Association, 2014). It is estimated that around 55 million people are currently affected by dementia worldwide, and this number is expected to triple by 2050 (WHO, 2019).

The most frequent forms of dementia observed in the elderly are Alzheimer's disease, which may contribute to 50-70% of cases, and vascular dementia, which accounts for 15-30% of cases (Alzheimer's Association, 2014). Alzheimer's disease (AD) is believed to be caused by the build-up of two proteins, amyloid and tau, which accumulate around brain cells and form deposit known as "plaques", in the case of amyloid, and "tangles" within brain cells, in the case of tau (Breijyeh & Karaman, 2020). These build-ups of protein cause the loss of brain cells, and may mediate a decrease in the levels of the acetylcholine neurotransmitter, involved in interneuron signal transmission, although both of these processes are not yet fully understood (NHS, 2021). The clinical presentation of Alzheimer's disease may be heterogeneous based on the brain areas affected by the presence of these plaques and tangles. The hippocampus appears to be frequently affected, causing the memory loss that is commonly associated with this disorder (NHS, 2021). Other typical symptoms of Alzheimer's disease include behavioural and personality changes such as apathy, irritability and depression (Alzheimer's Association, 2014). Vascular dementia, on the other hand, is caused by decreased blood flow to the brain, which damages neurons, eventually causing neuronal death and consequent loss of brain tissue (NHS, 2021). Reduced blood flow in the brain may be caused by the narrowing or obstruction of blood vessels in the brain, or by the temporary interruption of blood supply to the brain, like in the case of strokes (NHS, 2021). Again, vascular dementia may have varying clinical presentations based on the brain areas impacted by suboptimal blood flow.

As the prevalence of dementia rises, there are still limited treatment options available for degenerative disorders (Alzheimer's Association, 2014). Given the progressive nature of this disorder, the lack of available treatment raises concerns for the adequate management of patients with severe cognitive deficits, who may become functionally dependent in advanced stages of the disorder. In later stages of dementia, people with this disorder may indeed become confused, agitated, lose their awareness of time and place, self-neglect or engage in aggressive behaviour. When such complex clinical presentations of dementia arise, people affected by this disorder become fully reliant on carers, and might need to be admitted to long-term care facilities.

The primary risk factor for dementia is age: older people are more like to develop this disorder, which affects 7.1 % of people over the age of 65 (Alzheimer's Association, 2014). It follows that as populations grow older, the prevalence of dementia will continue to increase, while well-established treatment options are unavailable. As discussed to this point, this has significant implications for the quality of life of those affected. It also represents a significant challenge for healthcare, with considerable economic and societal costs. As life expectancy continues to increase, there is a growing sense of urgency for Governments and public health authorities to mitigate the pressure caused by dementia on healthcare systems.

Successful ageing and risk factors for dementia

Although age is a significant risk factor for dementia, this disorder is not a direct consequence of normal ageing. There is evidence showing that cognitive outcomes are heterogeneous between individuals (Harada et al., 2013). Although about 60% of the variability in cognitive profiles across individuals can be attributed to genetic determinants (McLearn et al., 1997), there are several environmental factors which appear to have an influence on the extent of cognitive decline in older age, and have been correlated to "successful ageing". This concept has been explained by different theoretical frameworks. For instance, the cognitive reserve hypothesis posits that some individuals may have a superior cognitive reserve, allowing them to recruit particularly efficient networks in order to maximise cognitive performance (Stern, 2002; Nucci et al., 2012). This reserve would be influenced by environmental factors such as education, social engagement, socioeconomic status, and is thought to protect the brain against the structural and functional changes associated with ageing. Similarly, the lifestyle-cognition hypothesis (Fratiglioni et al., 2004, Marioni et al., 2012) holds that an active lifestyle characterised by cognitively stimulating activities, social engagement and physical activities such as cardiovascular exercise, may contribute to prevent cognition from declining in older age and reduce the chances of developing dementia, as opposed to a less engaging lifestyle.

The notion that some environmental aspects of general lifestyle may be related to better cognitive outcomes can be used to plan and implement strategies aimed at preventing dementia onset. Building on this concept, disease prevention programmes can be developed, which aim at preserving cognitive functioning by targeting modifiable aspects of general lifestyle. In order to plan and develop these programmes, public health research has been increasingly focusing on investigating *risk factors for dementia*. Risk factors for dementia are aspects of genetic background, lifestyle and environment which may increase the risk of developing dementia, or negatively influence cognitive trajectories in older adults. While dementia has well-documented economic, societal and personal implications for those affected, even moderate rates of cognitive decline may lead to functional dependence and reduce quality of life of those who experience it. Therefore, investigating risk factors for dementia which influence cognitive trajectories is of great importance, as the application of this research to public policy making may benefit a wider spectrum of the population.

Risk factors do not represent the cause of a disease: rather, the exposure to risk factors determines an increased risk that this disease will develop. When discussing risk factors for dementia, it is important to distinguish between *modifiable* and *non-modifiable* risk factors. While the former can be modified and therefore targeted through interventions, the latter are permanent characteristics that cannot be changed. The most common non-modifiable risk factors for dementia are age, gender and genetic determinants (Rolandi et al., 2020). Particularly worthy of mention is APOE4, a form of the apolipoprotein E gene, which is involved in the metabolism of fats in humans. APOE4 has been identified as one of the strongest non-modifiable risk factors for developing Alzheimer's disease. This form of the APOE gene has been shown to increase the risk of developing Alzheimer's disease for those who carry it by 2 to 3 times, possibly by enabling vascular determinants of cognitive impairment such as the hardening of arteries supplying blood to the brain (Michaelson, 2014).

On the other hand, modifiable risk factors for dementia are aspects of lifestyle that can be modified throughout the lifespan in order to prevent dementia onset. Emerging research shows that nearly 35% of dementia cases could be prevented by targeting nine modifiable risk factors for dementia, which have been established as low educational attainment, hearing loss, smoking, hypertension and obesity in midlife, depression, diabetes, physical inactivity, and reduced social engagement in late-life (Livingston et al., 2017, Livingston et al., 2020; Rolandi et al., 2020). Barnes and Yaffe (2011) reviewed data from a systematic review conducted by the National Institutes of Health in 2010 (Daviglus et al., 2010). By analysing population risks calculated from the prevalence of specific risk factors and their correlation with the outcome variable, the researcher's review established that an improvement of 10-25% across only seven of the abovementioned risk factors (hypertension, obesity, diabetes, depression, low educational attainment, physical inactivity and smoking) could prevent up to 1.1-3 million cases of Alzheimer's disease globally (Barnes & Yaffe, 2011; Chen et al., 2014). These findings highlight the potential that modifiable risk factors hold against cognitive decline, and reinforce the urge for public health research to further investigate their potential. Preventable medical risk factors such as heart disease, traumatic brain injury, delirium, and stroke have also been shown to increase dementia risk (Rolandi et al., 2020).

Physical fitness and cognitive trajectories

The current investigation will focus on investigating the preventive effect of one such modifiable risk factor, physical fitness, against cognitive decline and dementia onset.

In the last twenty years, physical fitness has gained increasing attention because of its preventive potential against cognitive decline (Barness & Yaffe, 2011). There is compelling evidence that leading a physically active lifestyle characterised by greater levels of physical fitness promotes improved overall health outcomes, and holds protective effects against cognitive deterioration in older adults, as well as acting as a preventive factor against dementia (Mandolesi et al., 2018). For instance, Yaffe et al. (2009) showed greater levels of self-reported weekly moderate/vigorous exercise levels to be correlated with reduced cognitive decline over time. Hotting and Roder's (2013) review of the literature showed that physical exercise interventions may increase neuroplasticity and preserve cognitive ability. Cross-sectional associations between high levels of physical fitness and optimal cognition in older adults have been widely documented (Demnitz et al., 2016; Falck et al., 2017). The common cause theory, advanced by Christensen and colleagues (2001) builds on such cross-sectional associations between physical and cognitive performance. The common cause theory proposes that age-related changes in motor and cognitive performance may be due to a common unifying mechanism which may account for changes across the two different domains, and provide a common cause for ageing. Christensen and colleagues' theory would therefore explain why declines in physical and cognitive performance may be associated, and has since paved the way for research looking at correlations in the decline of physical functioning and cognition. While the nature of the common cause is still uncertain, the authors proposed that the factor underlying the impaired cognitive and non-cognitive processes observed in their analysis may be a marker age-related biological changes such as white matter changes (Christensen et al., 2001). Alternatively, they propose that the individual's conscious understanding may be the underlying common cause, due to the fact that the measures loading on this common factor all required the individual's understanding of instructions in order to be completed (Christensen et al., 2001).

Evidence from prospective research further elucidates cross-sectional findings by showing that higher levels of physical fitness are associated with reduced risk of developing cognitive impairment or receiving a dementia diagnosis in later life (Hamer et al., 2009; Sofi et al., 2010; Clouston et al., 2013; Blondell et al., 2014). Longitudinal evidence shows that the risk of developing cognitive impairment is as high as 3.5 times greater in sedentary older adults as opposed to older adults who walk more than 3 miles a day (Van Gelder et al., 2004). Longitudinal evidence strengthens the argument that physical fitness may have a protective role against cognitive decline or reduce the degree of cognitive deterioration over time. This preventive action is believed to occur by the enablement of various brain mechanisms. Firstly, physical activity is thought to facilitate plastic structural brain processes, such as an increase in the grey matter volume of frontal regions of the brain (Colcombe et al., 2006) and the hippocampus (Erickson et al., 2011). Indeed, longitudinal neuroimaging evidence has shown that adults with higher baseline levels of physical activity have lower frontal lobe atrophy progression compared to less physically active adults over an 8-year period follow-up (Yuki et al., 2012). In addition, physical activity facilitates the release of brain-derived neurotrophic factor (BDNF), a protein that regulates and promotes the growth of neurons in the adult brain (Kirk- Sanchez & McGough, 2014). Lastly, optimal levels of physical activity have also been linked to an increase in cerebral blood flow, which results in enhanced cognitive functioning and improved cardiovascular health (Ainslie et al., 2008; Ogoh et al., 2014).

The WHO (2020) recommends that adults should perform at least 150–300 minutes of moderate aerobic physical activity, or 75–150 minutes of vigorous aerobic physical activity throughout the week. Recent survey data shows that the proportion of the UK population meeting these aerobic activity guidelines decreases steeply with age (Scholes, 2017). At the same time, the proportion of survey respondents classed as inactive (i.e., performing less than 30 minutes of moderate to vigorous physically activity) increased with age, with 58% of women and 47% of men aged 75 and over being categorised as inactive (Scholes, 2017). Considering the well-documented health benefits of a physically active lifestyle both on cognition and general wellbeing, encouraging older adults to replace sedentary time with physical activity of any intensity through disease prevention programmes could contribute towards promoting healthier ageing outcomes.

Establishing temporal associations: the significance of longitudinal evidence

Investigating the risk factors that may determine successful ageing is highly significant as life expectancy increases worldwide. In order to fully comprehend the extent to which risk factors may play a role in preventing dementia, a life-course, prospective approach should be adopted, where risk factors are evaluated before the onset of dementia takes place (Singh-Manoux & Kivimaki, 2010). Prospective studies looking at participants without evidence of cognitive impairment at baseline are an optimal means of investigation for this purpose, as they grant the evaluation of the impact of the risk factor prior to the presence of any observable cognitive deficit that may be attributable to dementia. Longitudinal trajectories showing reduced risk of cognitive decline in individuals who were physically active in earlier life will then allow to establish a directionality of this effect, which supports the hypothesis that lower exposure to physical activity during earlier years might be a risk factor for dementia onset in later life. The accurate study of cognitive trajectories requires that the same group of individuals are assessed over time, in order to clearly define the relation between risk factors and outcomes of interest (Hofer & Sliwinski, 2001). Indeed, this approach allows researchers to evaluate important early stages of exposure to modifiable risk factors, that cannot be encompassed by cross-sectional correlational analysis in late life. Additionally, the utilisation of a longitudinal approach allows researchers to obtain repeated assessments of risk factors and cognition, allowing for a measurement of the variables of interest that is sensitive to change, as well as capable to assess their long-term relation (Singh-Manoux & Kivimaki, 2010).

Most importantly, prospective studies investigating this risk factor for dementia help to establish the evidence base for a physical function-cognition association, which can positively contribute in guiding public health policy making towards the development of disease prevention programmes encouraging adults to engage in physical activity starting from midlife, given its well-documented beneficial role on cognition. In light of clinical evidence showing that dementia may develop over many years, and remain latent for as long as 20 years before its observable onset (Alzheimer's Association, 2009), the adoption of a longitudinal perspective to investigate risk factors for dementia acquires further relevance.

Performance-based measures of physical fitness in ageing research

In spite of the importance of prospective longitudinal studies investigating the relation between physical fitness and cognition, there is still a relatively small number of studies targeting this specific association, with great heterogeneity in the outcome measures employed to assess both variables of interest. Here, we will discuss the use of commonly used self-report measures of physical activity levels in ageing research, and highlight the importance of adopting an objective approach, where this variable is assessed through performance-based measures.

Physical fitness levels are most commonly assessed using self-report questionnaires in epidemiological research, mostly in light of their practicality and reduced costs (Ainsworth et al., 2012; Innerd et al., 2015). Indeed, the majority of the literature investigating the prospective association between fitness and cognition employs subjective measures, where participants are asked to self-report their levels of engagement in physical activity (Hamer et al., 2009; Sofi et al., 2010; Blondell et al., 2014). Specifically, self-report measures may either assess self-rated physical activity, or quantified self-reports of this variable (Watkinson et al., 2010). Self-rated measures ask participants to evaluate their levels of engagement in physical activity, by choosing one response from a scale of options. In other words, they provide a single score which encompasses their overall rate of physical activity (Watkinson et al., 2010). In contrast, self-report measures ask respondents to quantify their levels of physical activity over a specific timespan through the use of questionnaires. This is often done by specifying the number of minutes/hours spent doing physical activity per week, month or year (Watkinson et al., 2010). Responses may then be compared to physical activity guidelines in order to establish adherence rates, and participants may then be accordingly grouped based on different levels of activity. The most widely used self-report measures of physical activity include the Physical Activity Scale for the Elderly (PASE), the Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH), the Baecke Physical Activity Questionnaire (BPAQ), and the International Physical Activity Questionnaire (IPAQ) (Healey et al., 2020). All of these measures are self-completed, and most of them enquire about activity levels in the 7 days prior to completing the questionnaire, aside from the PASE, which has a recall period of “the past few months”. While the BPAQ, PASE and SQAUSH focus mostly on sport, occupational, leisure time and household physical activities in a typical week, the IPAQ specifically asks about time spent on moderate-to-vigorous intensity activities, walking and sitting (Healey et al., 2020). All of these questionnaires are characterised by low participant and administrative burden, as they are short to administer (duration ranging from 5-15 minutes), and easy to score.

There are a number of issues with the use of self-report questionnaires in ageing research, especially in light of cognitive characteristics of the elderly population, which will be further discussed in chapter 2 below. Namely, questionnaires need to be designed in a way that minimises recall bias, and that effectively evaluates activities that are significant for the elderly population, in light of their cognitive deficits and functional limitations (Shepard, 2003; Innerd et al., 2015). When self-reporting physical activity, participants may also under or over-estimate their physical activity engagement, which may yield an inaccurate depiction of their physical status, and confound the correlation between physical fitness and cognition.

The paucity of objective measures of physical activity has been stressed by current epidemiological research. In a recent systematic review of prospective studies investigating longitudinal associations between physical fitness levels and cognitive decline in later life, the authors retrieved only two articles measuring physical function objectively, and consequently highlighted the need for increased employment of objective measures of physical fitness in longitudinal cohort studies (Blondell and colleagues, 2014). The employment of performance-based measures when assessing physical fitness ensures objective measurement of the variable of interest, as well as a comprehensive assessment of a remarkably multifaceted concept. Indeed, physical fitness involves various factors which are related to both the skill and the health status of the individual: it can be defined as a complex construct composed by different aspects, amongst which the primary ones are muscular strength, muscular endurance, and cardiorespiratory endurance (Caspersen et al., 1985). These are respectively the amount of force a muscle can produce with a single effort, the ability to sustain muscle contraction for a relatively long period of time, and the ability to supply oxygen and other essential nutrients to working muscles. Research focusing on the associations between physical functioning and cognition should include objective measures that encompass at least one of these components of physical fitness, in order to provide an accurate depiction of an individual's physical fitness status. Objective measures of physical fitness have been available and employed in research for nearly 40 years; they primarily include grip strength as measured with a dynamometer, walking speed, standing balance and chair rises (Clouston et al., 2013; Blondell et al., 2014). These measures offer a comprehensive picture of the physical status of the individual, which is not subject to self-report recall bias, or under and overestimation biases, as well as being highly reliable and reproducible (Ainsworth et al., 2012). Further, objective measures of physical fitness can provide a continuous marker of physical function rather than dichotomised outcomes expressing compliance to recommended guidelines (or lack thereof), such as in the case of self-report measures. Innovative methods to objectively assess physical function are becoming increasingly popular in epidemiological research, partially as a practical response to the self-report bias present with questionnaires.

Instruments such as accelerometers and pedometers are easily accessible, cheap and reliable instruments to objectively measure physical function (Sesso, 2007; Parker et al., 2008). While pedometers objectively assess the number of steps taken, which can be used to estimate overall fitness levels, accelerometers also measure accelerations and decelerations, which provide a useful indication of the intensity of different activities. These wearable devices are gaining increasing popularity in ageing research on physical fitness. Whilst traditional wearable trackers used to be bulky, and to require manual recording of data, recent advances in technology have allowed for small, wrist-worn devices such as smart watches and wrist bands to transmit data wirelessly to mobile applications, making their use more comfortable and efficient. Recent studies looking at the acceptability and adherence rates to wrist-worn activity trackers in adults aged 65 or older shows average adherence rates to wearing the device of 95%, with participants reporting high comfortability (4.63/5) with the daily use of wearable activity trackers (Paolillo et al., 2022), average rates of acceptability of 80% (Zhang et al., 2022), and usability rates of 93% (Domingos et al., 2022). These figures suggest high wearability of wrist-worn activity trackers, and indicates their suitability to be used in research with older adults.

In the next chapter, a systematic review of the literature will be conducted, to summarise the evidence from prospective cohort studies investigating the longitudinal association between performance-based measures of physical fitness and cognition in older adults in the absence of cognitive impairment at baseline. As discussed above, the majority of the epidemiological studies investigating this association employs self-report measures of physical activity. Further to this, the already existing systematic reviews addressing the topic tend to focus on the protective role of physical fitness against the risk of developing dementia/Alzheimer's disease over time, as measured through diagnostic criteria (Hamer et al., 2009). This approach may not be fully representative of the different degrees of cognitive decline experienced by the older population, as individuals with a low level of cognitive impairment may not reach a level of decline that is compatible with a diagnosis of dementia but still exhibit cognitive changes.

Chapter 2: Longitudinal association between performance-based physical fitness measures and cognitive decline: a systematic review of prospective studies

Introduction

The potential protective role that physical fitness may hold against cognitive decline has been supported by two recent reviews establishing it as one of the top seven modifiable risk factors for cognitive decline in older adults (Barnes & Yaffe, 2011; Deckers et al., 2015). Although this association is increasingly well-documented as it gains further attention in ageing research, the majority of existing reviews on the topic examines evidence from studies using subjective measures of physical activity, and there is great heterogeneity in the predictive and outcome measures employed. The paucity of longitudinal studies adopting an objective approach towards measuring the protective effect of physical function on cognition has been highlighted by a recent review on the topic (Blondell et al., 2014). This chapter will systematically review the evidence from prospective studies investigating the association between performance-based measures of physical fitness and cognition in adults aged 50 or over, with no diagnosis of neurodegenerative disease at baseline. The primary objective of this review is to investigate the preventive potential of objectively-assessed physical fitness against cognitive decline. As discussed above, this review aims to fill gaps in previous research, by examining research employing performance-based measures of physical fitness. This review will exclusively include studies using objective measures such as handgrip strength, accelerometer-based or performance-based walking speed, standing balance, chair rises, cardiovascular fitness measures and composite performance-based measures, characterised by a combination of physical tasks. These measures capture a reliable picture of the individual's physical capability, which offers a good indication of their ability to perform daily life activities. Further, this review will consider outcomes of cognitive decline, as well as diagnoses of dementia or Alzheimer's disease. This is in order to provide a comprehensive representation of cases of cognitive impairment that may not be severe enough to yield a diagnosis of dementia.

Existing reviews

Prior to starting the search process, we reviewed the already existing and recently conducted systematic reviews looking at the prospective association between physical fitness and cognition in older adults. It must be noted that the majority of the reviews on the topic looked at the effects of short-term physical exercise interventions on cognition in older adults (Busse et al., 2009; Gheysen et al., 2018), while the reviews focusing on our association of interest, i.e. the prospective associations between the variables under investigation as the primary outcome, were not numerous. Four recent systematic reviews targeting this association were identified. Hamer et al. (2009) reviewed 16 prospective cohort studies investigating the association between baseline physical activity levels and risk of neurodegenerative disease. Findings from this review suggest that greater physical activity levels are associated with reduced risk of dementia. Notably, all of the studies included assessed physical activity through self-reported questionnaires, or at times they simply provided a categorical "yes/no exercise" answer. Sofi and colleagues' (2010) systematic review and meta-analysis reviewed 15 prospective cohort studies looking at the association between physical activity and risk of cognitive decline in older adults. Their meta-analysis showed that higher self-reported levels of physical activity were associated with a 38% lower risk of cognitive decline. Again, all of the studies included in this review involved a self-reported assessment of physical activity. Blondell and colleagues (2014) conducted a review on 47 prospective cohort studies, with 21 looking at the association between physical activity and cognitive decline, and 26 investigating the correlation between physical activity and dementia risk. This review showed that greater baseline levels of physical activity were correlated with reduced risk of developing dementia or cognitive decline at follow up. Having only retrieved two articles which objectively assessed physical fitness, Blondell and colleagues indicated the need to expand the use of performance-based measures of physical fitness in ageing research, which we will aim to address in the current review. The outcomes investigated by Blondell and colleagues' study included both cognitive decline and dementia risk: this approach will also be followed in our review. The self-reported measures that were used in these reviews range from simple self-completed questionnaires enquiring on frequency and duration of different types of physical activities (walking, cycling, gardening etc.) which are then converted into a total weekly measure of physical activity in minutes or hours, to questionnaires administered by an interviewer enquiring on activity levels, on the frequency of different physical activities over different time spans, or on the average distance walked per day (Blondell et al., 2014). Other studies either simply asked about exercise or sports participation, and a yes/no answer defined participants as physically active or inactive, or enquired about the number of days of activity per week, self-reported walking distance, or self-reported levels of intense activities in a week (Hamer et al., 2009). Studies included in Sofi et al.'s (2010) review included studies which mostly used questionnaires enquiring about hours of exercise per week, or enquired on frequency and intensity of exercises. Validated questionnaires included the Zutphen Physical Activity Questionnaire, which defines a kilocalorie summary score from different types of activities such as walking or cycling, and the IPAQ (Blondell et al., 2014).

Clouston and colleagues' (2013) recent investigation reviewed 36 studies looking at the dynamic relation between changes in performance-based measures of physical fitness and changes in cognition in adults aged 40 years or older. Clouston et al.'s findings showed that results depended on measurement types, with changes in grip strength and walking speed most significantly predicting cognitive outcomes. While adopting an objective approach towards the measurement physical fitness, which partially addresses the issues pointed out by Blondell and colleagues (2014), this review looked at change-to-change associations between the variables of interest as the primary outcome, which excludes a significant number of studies, as most studies do not assess changes in physical function, but only baseline levels of this variable. Our review will focus on baseline physical function versus later cognitive decline or dementia diagnosis**.**

Since the above reviews were conducted, devices such as dynamometers and accelerometers have become more widely available, leading to an increased number of longitudinal studies using objective measures of physical fitness. Therefore, the present review will expand on existing evidence and provide an update of Clouston and colleagues' work. The current review will focus on prospective associations between baseline physical function in older adulthood (50 years or older) and risk of successive cognitive decline or dementia risk. While some of the reviews outlined above focus exclusively on the protective role of physical fitness against dementia risk as assessed through diagnostic outcomes (Hamer et al., 2009), or solely on cognitive decline as measured mostly through the MMSE (Sofi et al., 2010), our review will include both outcomes (separated over two chapters). This approach will ensure that the evidence reviewed is representative of the differential degrees of cognitive decline experienced in older age, and that it includes participants with a level of cognitive impairment which is not significant enough so as to yield a dementia diagnosis. Further, we will consider cognitive outcomes other than the MMSE, in order to comprehensively inspect cognitive changes across different domains of cognition. Our review fills the gap identified by Blondell et al. (2014) and will focus on the association between baseline levels of physical function and risk of cognitive decline or dementia. The reasoning behind this approach is the aim to investigate the protective power of earlier life physical performance on later life cognition, without considering changing trajectories in physical function, which may be influenced by confounding variables.

Further, studies looking at change-to-change trajectories are still relatively low in frequency, therefore limiting the search to this specific association may exclude valuable evidence.

Methods

Search strategy

This review followed the guidelines from the PRISMA statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher et al., 2009). A computerised literature search of relevant studies was conducted on the following electronic databases: MedLine, PsycINFO and the Cochrane Library of Systematic Reviews. The search was continuously updated from January 2021 to May 2021. The following search strategy was used, with the relevant search terms being used both as title words, keywords and as Medical Subject Headings (MeSH): (“physical fitness" OR "physical activity" OR “physical capability” OR "physical exercise" OR “physical function*” OR exercise OR fitness) AND (cognition OR "cognitive function*" OR "cognitive impairment" OR dementia OR “cognitive decline”) AND (prospective OR cohort OR longitudinal AND "older adults" OR ageing OR "older people" OR 50+). Only studies in English language were included in the review. The database search strategy was supplemented by searching the references of relevant articles, as well as inspecting references of already existing reviews with similar inclusion criteria and objectives as the current review.

Eligibility criteria

Studies looking at the prospective association between performance-based measures of physical function and cognitive impairment over time were included if they met the following inclusion criteria: 1) a prospective or retrospective cohort study design, 2) a sample of community-dwelling adults aged 50 years or older and without diagnosis of neurodegenerative disorder or MCI at baseline, 3) the relation between physical function and cognitive decline as the primary or secondary object of the investigation, 4) included at least one objective measurement of physical function at baseline and a measure of cognitive decline and 5) reported a statistical index of association between physical fitness at baseline and cognitive changes at follow-up examination. Studies were excluded if they: 1) were cross-sectional studies without a prospective association, or intervention studies where the variable of interest was manipulated, 2) did not include at least one objective physical fitness measure and one cognitive outcome, 3) included participants who had a diagnosis of dementia or MCI at baseline, 4) included participants who were younger than 50 years of age, or 5) did not report the original data, for instance systematic reviews.

Study selection and data collection

An independent reviewer assessed the articles retrieved by the search to assess eligibility for inclusion. The search retrieved a total of 2412 articles (1444 from MEDLINE and 968 from APA PsycINFO, no reviews were retrieved on Cochrane using this search strategy). The retrieved articles were transferred to Covidence, a reference management software where they were organised for study selection, which was performed initially for title, then for abstract and finally for full text screening. Here, duplicates (n= 47) were excluded by the software. Upon review of study titles, 1739 articles were excluded due to the topic being irrelevant for the aim of the review. Abstracts were reviewed for the remaining articles, and a total of 162 articles were identified as suitable for full-text review. Of these, 117 further studies were excluded, with the majority of the excluded articles employing self-reported measures of physical fitness (n= 67), and the remaining number of studies employing an inappropriate study design (Cross-sectional = 23, intervention studies = 17), or including participants with dementia or MCI at baseline (n = 10). Five further studies were included by scanning references of included articles, or reference lists of systematic reviews on the topic. The final sample consisted of 50 studies (See Figure 1 for PRISMA flow diagram). Data from these studies were extracted using a standardised form, including information on: first author and year of the study, country, study from which the data was obtained, participant demographics, length of study follow up, performance-based measures of physical function, cognitive or diagnostic outcomes, presence of covariates and statistical estimate of association between physical function and cognitive decline. The decision was made to divide the data extraction, meta-analyses, qualitative synthesis and discussion of the final 50 articles in two different reviews hereafter, based on whether they included a cognitive (N = 35), or a diagnostic (N= 15) outcome. The rationale for this was that of providing a better structure and having consistent outcome measures in the meta-analyses. The data extraction for studies employing a cognitive decline outcome is summarised in Table 1 below.

Flow diagram

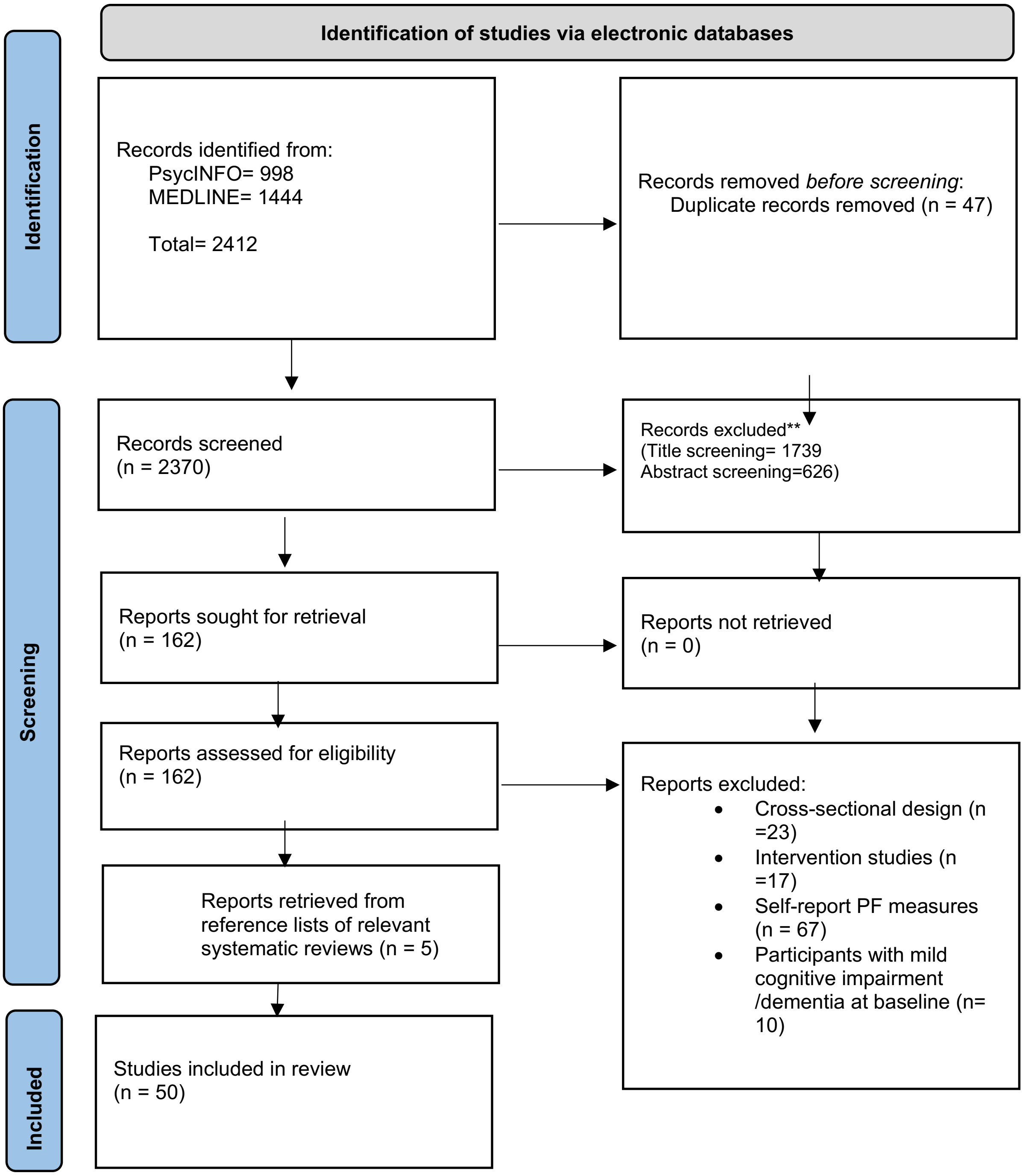

Figure 1: PRISMA flow diagram for the search of the systematic review

Data extraction

Table 1: Data extraction of relevant study variables, for studies with a cognitive decline outcome

First author, year	Country	Study	Participants	Baseline age (years)	Gender (Female %)	Length of follow-up	PF measure	Cognitive measure	Outcome
Abe 2017	Japan	The Kasama study	169	>65	47.30%	3 years	handgrip strength, balance test, sit-to-stand test, timed up and go test, walking speed, peg-moving task	composite score from "5 cog "cognitive testing battery evaluating attention, memory, visuospatial function, verbal fluency, and reasoning	changes in walking speed and peg-moving task were associated with cognitive decline
Alfaro-Acha, 2006	USA	The Hispanic Established Populations for the Epidemiologic Study of the Elderly (H-EPESE)	2160, 1303 at follow up	> 65	57%	7 years	handgrip strength	Mini Mental State Examination (MMSE)	participants with lowest handgrip strength at baseline had lower MMSE scores over time
Alfaro-Acha, 2007	USA	The Hispanic Established Populations for the Epidemiologic Study of the Elderly (H-EPESE)	2070, 1218 at follow up	> 65	57.5%	7 years	walking speed	MMSE	participants with lower walking time had a greater cognitive decline over time
Auyeung, 2011	China	Chinese University of Hong Kong Study	4000, 2737 at follow-up	> 65	45%	4 years	appendicular muscle mass (ASM), handgrip strength, sit-to-stand test, walking speed	MMSE	grip strength and chair-stand test were associated with a lower MMSE score at follow up in males, and grip strength was predictive of lower MMSE scores in females
Barnes, 2003	USA	Prospective cohort study of community-dwelling older adults living in Sonoma, California	349	≥ 59	49.20%	6 years	peak oxygen consumption (peak VO2)	MMSE, 3 tests of attention/executive function, 2 measures of verbal memory, and 2 tests of verbal fluency	participants with worse cardiorespiratory fitness at baseline experienced greater decline on the MMSE over 6 years
Best, 2016	USA	The Health, Aging, and Body Composition study	2678	> 70	52%	9 years	walking speed	Digit Symbol Substitution Test (DSST), modified mini-mental status examination (3MS)	baseline walking speed was a significant predictor of changes in later cognition as measured through 3MS

Boyle 2009	USA	The Rush Memory and Aging Project	970	> 54	75,15%	3.6 years	muscle strength, handgrip strength	MMSE and a composite cognitive score from 19 tests assessing different aspects of cognition	the rate of cognitive decline for participants with greater muscle strength was considerably slower than that of participants with lower strength
Buchman, 2007	USA	The Religious Orders Study	877	> 65	69.40%	5.7 years	handgrip strength	Alzheimer's disease diagnosis and cognitive decline assessed through a battery of 19 cognitive tests	a higher level of objectively-measured baseline frailty was associated with an increased rate of global cognitive decline
Buchman, 2012	USA	The Rush Memory and Aging Project	716	> 55	76%	4 years	accelerometer data	Alzheimer's disease diagnosis and cognitive decline assessed through a battery of 19 cognitive tests	levels of total daily physical activity were associated with the rate of global cognitive decline
Chen, 2020	Taiwan	Community-based project conducted in Hunei District, Kaohsiung (Taiwan)	285	> 65	55.40%	2 years	accelerometer data	subjective cognitive decline as measured via Ascertain Dementia 8-item Questionnaire	Participants in intermediate/ high walking groups showed reduced associations with cognitive decline compared to participants in low walking groups
Chou, 2019	Japan	The National Institute for Longevity Science – Longitudinal Study of Aging (NILS-LSA)	1063	> 60	49,1%	10 years	handgrip strength, walking speed	MMSE and DSST	lower gait speed and handgrip strength were associated with greater cognitive decline over time
Deshpande, 2009	Italy	The Invecchiare in Chianti Study	660, 584 at follow up	> 65	53%	3 years	walking speed	MMSE	lower performance in walking speed was a significant independent predictor of steeper decline of MMSE score over 3 years.
Gale, 2014	UK	The English Longitudinal Study of Ageing (ELSA)	2654	> 60	55.60%	6 years	walking speed	tests assessing verbal fluency, immediate/delayed verbal memory, and attention	participants with greater walking speed at baseline had less yearly decline on all cognitive domains
Halloway, 2017	USA	Prospective cohort study of community-dwelling latino adults aged 50 and older living in Chicago	59	> 50	78%	5 years	accelerometer data	neurocognitive tests assessing episodic memory, perceptual speed, and semantic memory	those who had less decline in accelerometer moderate–vigorous activity maintained semantic memory

Ho, 2001	Hong Kong	Data from participants registered to the Old Age Allowance (OAA) Scheme	988	> 70	48%	3 years	gait time	information/orientation part of the Clifton Assessment Procedure for the elderly (CAPE)	slower gait time was a predictor of cognitive impairment in both genders
Hsu, 2017	Taiwan	Longitudinal cohort study in a veterans' retirement community in southern Taiwan	249	> 80	Male 100%	1 year	handgrip strength, walking speed	MMSE	slow gait speed predicted rapid cognitive decline, while handgrip strength was not a significant predictor
Inzitari, 2007	USA	The Health, Aging and Body Composition Study	3075, 2,276 at follow up	> 70	53%	5 years	walking speed	DSST	those with lower gait speed had a significantly greater risk of experiencing a DSST 5-year decline
Katsumata, 2011	Japan	The Keys to Optimal Cognitive Aging Project (KOCOA)	192	> 80	73.40%	3 years	timed up and go test	MMSE	gait speed was associated with greater risk for global cognitive decline in the slower group in the TUG test
Kim, 2019	Korea	The Korean Longitudinal Study of Aging (KLoSA)	2,378	> 65	47.80%	8 years	handgrip strength	Korean Mini-Mental Status Evaluation [K-MMSE]	greater handgrip strength at baseline predicted MMSE scores positively over time
Kim, 2019	Korea	The Korean Longitudinal Study of Aging (KLoSA)	5,995	> 50	54.50%	8 years	handgrip strength	K-MMSE	For every 1 unit increase in handgrip strength, there was nearly 4% reduction in the risk of cognitive decline over time
Ku, 2017	Taiwan	Community-based project, conducted in Hunei District, Kaohsiung, Taiwan	285	> 65	54.40%	22 months	accelerometer data	Ascertain Dementia 8-item questionnaire	higher levels of sedentary behaviour were associated with an increased risk of cognitive decline at follow up
MacDonald, 2017	Canada	The Victoria Longitudinal Study	121	> 55	64%	25 years	walking speed, two markers of gait function derived from the GAITRite system (computerised walkway)	DSST, word recall, incidental and semantic memory tests	each additional second increase in timed walk was associated with a further decline in digit symbol performance accuracy and in the number of words successfully recalled.
Meunier, 2021	USA	The Cardiovascular Health Study	4,811	> 65	66%	6 years	balance test	3MS and DSST	participants with lower balance had a faster rate of cognitive decline
Mielke, 2013	USA	The Mayo Clinic Study of Aging	1,478	> 70	48,4%	4.11 years	walking speed	9 cognitive tests assessing memory, language, executive function, visuospatial	faster baseline walking speed was associated with reduced cognitive decline in all cognitive domains

								skills, and global cognitive scores.	
Muscari, 2018	Italy	The Pianoro study	405	> 65	51.80%	7 years	step test	MMSE	step test duration was associated with a worse performance in orientation, attention, calculation and language, and global cognitive decline
Okely, 2020	UK	The Lothian Birth Cohort 1936 (LBC1936)	1091	> 70	49.80%	9 years	Forced expiratory volume (FEV1), handgrip strength, walking speed	13 cognitive tests assessing 4 domains of verbal memory, processing speed, crystallised abilities and visuospatial ability	decline in verbal memory was preceded by declines in walking speed and grip strength
Ritchie, 2016	UK	The Lothian Birth Cohort 1936 (LBC1936)	1091	> 70	49.70%	6 years	Forced expiratory volume (FEV1), handgrip strength, walking speed	4 tests evaluating visuospatial reasoning and working memory, from which a fluid intelligence "g" factor was computed	baseline levels of physical functions did not predict cognitive decline
Sintjines, 2017	Netherland s	The Longitudinal Aging Study Amsterdam (LASA)	3102, 2545 at follow up	> 55	53, 4%	5-12 years	handgrip strength, walking speed	MMSE, Alphabet Coding Task-15, 15-Words test	slower walking speed and grip strength were associated with a steeper decline in cognition in adults aged 65-74; only slower gait speed in the 75-85 age group
Sintjines, 2017	Netherland s	Tthe Leiden 85-plus Study	599, 434 at follow up	> 85	53, 4%	5-12 years	handgrip strength, walking speed	MMSE, the Letter Digit Substitution Task (LDST), 12-Picture Learning test	weaker handgrip strength was associated with a steeper decline in global cognitive function.
Stubbs, 2017	Taiwan	Community-based project conducted in Hunei District, Kaohsiung (Taiwan)	274	> 65	54,4%	22 months	accelerometer data	Ascertain Dementia 8-item questionnaire	higher levels of light PA were associated with a reduced rate of decline in cognitive ability
Suzuki, 2017	Japan	Study conducted in residential areas surrounding the Tokyo Metropolitan Institute of Gerontology (TMIG)	496	> 65	55.80%	1 year	handgrip strength, timed up and go test, walking speed	Japanese version of the Montreal Cognitive Assessment (MoCA-J)	higher for lower performance in TUG and walking speed were associated with greater odds of decline vs improvement in MoCA-J performance at follow-up
Veronese, 2016	Italy	The Progetto Veneto Anziani (PRO.V.A.),	1249	> 65	59,5%	4.4 years	Short Physical Performance Battery (SPPB), balance test, walking speed, sit-	MMSE	participants with lower SPPB performance or slowest walking speed were more likely to develop cognitive decline

							to-stand test, 6minute walking test, handgrip strength, Leg strength		
Viscogliosi, 2017	Italy	Observational prospective study in a Geriatric outpatients center in Rome, Italy	104	> 70	56%	11.2 months	handgrip strength	Clinical Dementia Rating scale and the Clock Drawing Test (CDT)	an association was observed between baseline handgrip strength and 1-year cognitive decline
Wadsworth , 2020	USA	The Hispanic Established Populations for the Epidemiologic Study of the Elderly (H-EPESE)	1489	> 65	62.90%	9 yeas	balance test, walking speed, sit-to-stand test (composite score)	MMSE	walking speed and timed chair stands, but not balance were associated with greater cognitive decline
Zaninotto, 2018	UK	The English Longitudinal Study of Ageing (ELSA)	10626	> 50	54.40%	8 years	walking speed	battery of cognitive tests assessing 3 domains of memory, processing speed, cognitive function	gait speed was not related to the rate of change of any domain of cognitive function
Zhu, 2017	USA	The Reasons for Geographic and Racial Differences in Stroke (REGARDS) study	6435	> 60	55.30%	3 years	accelerometer data	Six-Item screener, letter fluency, animal fluency, word list learning, and MoCA	participants with higher levels of vigorous activity had less cognitive impairment over time

Characteristics of the studies

A total number of 35 studies investigating the association between physical function at baseline and later life cognitive decline was retrieved. The total number of participants from the studies included was 63928, with participants from all studies aged 50 years or older. The sample size of the studies included ranged from N= 59 to N = 10626. All of the studies included were prospective longitudinal cohort studies, with follow-up durations ranging from a minimum of 11.2 months to a maximum of 25 years. Most of the studies included were conducted in the USA (13 studies), followed by China (6 studies), Japan, Italy and the United Kingdom with 4 studies each, Korea (2 studies), Canada and Netherlands with 1 study each. All the studies included both genders, except for one (Hsu, 2017). The studies included in the review employed a variety of performance-based measures of physical function: 4 studies relied exclusively on handgrip strength, 9 studies employed walking speed exclusively as an indication of physical performance, 6 studies assessed daily step count as assessed through accelerometer wearing, one study used only tandem balance measures, and another one used solely the timed up and go test, and finally one study looked only at cardiovascular measures of physical function, namely peak oxygen consumption. The remainder of studies (13) used different combinations of the above measures. A total of 16 studies included handgrip strength, 20 studies used walking speed, 4 studies employed the sit-to-stand test, 4 studies used balance measures, 6 looked at accelerometer data, 3 studies used cardiovascular measures and 3 studies assessed the Timed Up and Go performance. The 35 studies included either a single neuropsychological test, the most common of which was the Mini Mental State Examination (MMSE), or a battery of cognitive tests to assess cognitive status. Most of the studies included covariates of the analyses, the most common of which were age, sex, education, co-morbid health conditions, BMI, depression, baseline cognition, and social engagement.

Results

Assessment of study quality and covariates

Quality assessment was conducted using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Cohort Studies (Moola et al., 2017). This tool is used to assess the quality of prospective and retrospective cohort studies by investigating information on measurement of the exposure to the variable of interest, identification and inclusion of confounding variables in statistical analyses, length of follow-up, validity of outcome measures and loss of follow-up rate. Additionally, this tool was modified by adding information on sample size, and by specifying relevant covariates that should be included in longitudinal statistical analyses. Based on whether studies met these criteria, a score ranging from 0 (corresponding to No or N/A), 1 (Unclear/Partially), or 2 (Yes) was attributed to each criterion. A total score was computed for each study, whereby a score of 16-20 identified high-quality studies, a score of 11-15 identified studies of fair quality and a score of 5-10 identified studies of poor quality. The mean quality of the studies included was 15.80, with scores ranging from 10 to 20 and with an overall fair to high level of the studies assessed. One study was defined as being of poor quality, 14 studies were of fair quality, and 20 were of high quality. The quality assessment has been summarised in Table 2 below (see p.37).

The studies included in this review adjusted their analyses according to covariates that can be divided in 6 categories, which include: (1) demographic factors such as age and gender, (2) social factors such as socioeconomic status and education, (3) lifestyle factors such as smoking, diet, or alcohol intake (4) comorbidities such as cardiovascular disease, hypertension or diabetes, (5) health status, such as depression, anxiety symptoms, body mass index, or cognitive status and (6) genetic factors such as *APOE4* status. While most of the studies included in our review reported that covariates did not influence the significance of the association between physical fitness and cognitive decline, some of the studies reported that the covariates either attenuated or explained the association completely. Of interest, Auyeung et al. (2011) found that demographic factors explained the association between baseline physical function and cognitive decline, while Barnes et al. (2003) found that demographic and health-related outcomes attenuated the same association. Viscogliosi and colleagues (2017) found that white matter hyperintensities attenuated the association between physical function and cognitive decline. Comorbidities and chronic illness were also found to explain the association between physical function and cognitive decline (Gale et al., 2014). Deshpande et al. (2009) found that the association between walking speed and cognitive decline was not significant after controlling for baseline MMSE scores, demographic and health-related confounders.

Table 2: Assessment of study quality

First author, year	Were the two groups similar and recruited from the same population?				Were the exposures measured similarly to assign people to both exposed and unexposed				Was the exposure measured in a valid and reliable way?				Were confounding factors identified? (at least age, gender, education,			
	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A
Abe 2017				0				0	2				2			
Alfaro-Acha, 2006				0				0	2				2			
Alfaro-Acha, 2007				0				0			1		2			
Auyeung, 2011				0				0	2				2			
Barnes, 2003				0				0	2				2			
Best, 2016				0				0	2				2			
Boyle 2009				0				0	2				2			
Buchman, 2007				0				0	2				2			
Buchman, 2012				0				0			1		2			
Chen, 2020				0				0	2				2			
Chou, 2019				0				0	2				2			
Deshpande, 2009				0				0	2				2			
Gale, 2014				0				0			1		2			
Halloway, 2017				0				0			1		2			
Ho, 2001				0				0	2				2			
Hsu, 2017				0				0	2				2			
Inzitari, 2007				0				0			1		2			
Katsumata, 2011				0				0			1		2			
Kim, 2019				0				0	2				2			
Kim, 2019				0				0	2				2			
Ku, 2017				0				0	2				2			
MacDonald, 2017				0				0	2						1	
Meunier, 2021				0				0	2				2			
Mielke, 2013				0				0	2				2			
Muscari, 2018				0				0			1		2			
Okely, 2020				0				0	2				2			
Ritchie, 2016				0				0	2				2			
Sintjines, 2017				0				0	2				2			
Stubbs, 2017				0				0	2				2			
Suzuki, 2017				0				0	2				2			
Veronese, 2016				0				0	2				2			
Viscogliosi, 2017				0				0	2				2			
Wadsworth, 2020				0				0	2				2			
Zaninotto, 2018				0				0	2				2			
Zhu, 2017				0				0			1		2			

First author, year	Were strategies to deal with confounding factors stated?				Were the groups/participants free of the outcome at the start of the study?				Were the outcomes measured in a valid and reliable way?				Was the follow up time reported and long enough for outcomes to occur?			
	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A
Abe 2017	2				2				2						1	
Alfaro-Acha, 2006	2				2				2				2			
Alfaro-Acha, 2007	2				2				2				2			
Auyeung, 2011	2				2				2						1	
Barnes, 2003	2				2				2				2			
Best, 2016	2						1		2				2			
Boyle 2009	2				2				2						1	
Buchman, 2007	2				2				2				2			
Buchman, 2012	2				2				2						**1**	
Chen, 2020	2					0				0					1	
Chou, 2019	2				2				2				2			
Deshpande, 2009	2						1		2				2			
Gale, 2014	2				2				2				2			
Halloway, 2017	2				2				2				2			
Ho, 2001	2				2				2						1	
Hsu, 2017	2						1		2					0		
Inzitari, 2007	2						1		2				2			
Katsumata, 2011	2						1		2						1	
Kim, 2019	2				2				2				2			
Kim, 2019	2						1		2				2			
Ku, 2017	2						1			0				0		
MacDonald, 2017			1		2				2				2			
Meunier, 2021	2				2				2				2			
Mielke, 2013	2				2				2						1	
Muscari, 2018	2						1		2				2			
Okely, 2020	2						1		2					0		
Ritchie, 2016	2						1		2				2			
Sintjines, 2017	2					0			2				2			
Stubbs, 2017	2					0				0				0		
Suzuki, 2017	2						1		2					0		
Veronese, 2016	2						1		2						1	
Viscogliosi, 2017	2				2				2					0		
Wadsworth, 2020	2						1		2				2			
Zaninotto, 2018	2						1		2				2			
Zhu, 2017	2						1		2						1	

First author, year	Was follow up complete, and if not, were the reasons to loss to follow up described?				Were strategies to address incomplete follow up utilized?				Was appropriate statistical analysis used?				Was the sample size appropriate?				Score:
	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	
Abe 2017		0				0			2						1		14
Alfaro-Acha, 2006	2				2				2				2				20
Alfaro-Acha, 2007	2				2				2				2				19
Auyeung, 2011			1			0			2				2				16
Barnes, 2003			1			0			2						1		16
Best, 2016	2				2				2				2				19
Boyle 2009	2							0	2						1		16
Buchman, 2007	2							0	2						1		17
Buchman, 2012	2							0	2						1		15
Chen, 2020	2							0	2					0			11
Chou, 2019	2				2				2						1		19
Deshpande, 2009	2					0			2						1		16
Gale, 2014		0				0			2				2				15
Halloway, 2017	2					0			2					0			15
Ho, 2001	2				2				2						1		18
Hsu, 2017	2							0	2					0			13
Inzitari, 2007	2				2				2				2				18
Katsumata, 2011			1			0			2					0			12
Kim, 2019	2				2				2				2				20
Kim, 2019		0				0			2				2				15
Ku, 2017			1			0			2					0			10
MacDonald, 2017		0				0			2					0			12
Meunier, 2021	2							0	2				2				18
Mielke, 2013	2							0	2				2				17
Muscari, 2018	2							0	2					0			14
Okely, 2020	2				2				2						1		16
Ritchie, 2016			1			0			2						1		15
Sintjines, 2017	2				2				2					0			16
Stubbs, 2017	2				2				2					0			12
Suzuki, 2017	2				2				2					0			15
Veronese, 2016	2				2				2				2				18
Viscogliosi, 2017	2				2				2					0			16
Wadsworth, 2020			1				1		2				2				17
Zaninotto, 2018			1		2				2					2			18
Zhu, 2017			1				1		2				2				15

Y= Yes, N= No, U/P= Unclear/Partially, N/A= Not applicable

Statistical analysis

Several meta-analyses were performed to investigate the association between baseline physical function and cognitive decline, where pooled mean estimates were computed using the review and meta-analysis software RevMan review manager 5.4.1. Given the heterogeneity of follow up periods and sample sizes, random-effect models were employed to calculate pooled effect sizes from either β coefficients and Standard errors, or Odds/Rate/Hazard Ratios and 95% confidence intervals. We adopted the strategy of grouping studies into different meta-analyses according to common predictors, outcomes and statistical coefficients of association, in order to mitigate the heterogeneity of the analysis as much as possible. Five objective markers of physical function were identified, whereby a sufficient number of studies for each predictor granted the possibility to conduct a meta-analysis using such variable as a predictor. These predictors were defined as walking speed, handgrip strength, sit-to-stand test, balance and accelerometer-assessed physical activity. The predominant outcome measures were either cognitive decline over time as assessed through *the Mini-Mental State Examination (MMSE)/ Modified Mini-Mental State (3MS)*, the *Digit Symbol Substitution Test (DSST)* or the *Ascertain Dementia 8-item Questionnaire.* Publication bias was examined using funnel plots, which were produced using RevMan 5.4.1. The Cochrane recommendations systematic reviews (Higgins et al., 2019) indicate that approximately 10 studies should be included in order to assess whether funnel plot asymmetries are driven by study characteristics rather than by chance. However, given the heterogeneity of the outcome measures considered, which resulted in a relatively small number of studies included in each meta-analysis, the present meta-analyses adopted a more flexible approach, where at least 5 studies were deemed necessary to assess risk of bias. Heterogeneity between the studies included in each meta-analysis was assessed via the I^2 statistic and the associated p value provided by RevMan 5.4.1, where significant values between 25% and 50% indicated low heterogeneity, scores between 50% and 75% indicated moderate heterogeneity between studies, and values greater than 75% suggested high heterogeneity between studies. Where heterogeneity was high, the analysis was followed by sensitivity analysis which excluded individual studies from each meta-analysis, to investigate the influence of single studies on mean estimates.

Meta-analysis

Walking speed and Mini-Mental State Examination

Eight studies were identified looking at the association between physical function as measured through walking speed, and cognition as measured through the MMSE. These papers mostly used regression models looking at the predictive role of baseline walking speed on either the absence/presence of cognitive decline, or on MMSE scores as a continuous variable. Of these studies, Sintjines and colleagues (2017) used crossed-lag panel models to analyse data from two different cohort studies. These models yield statistical coefficients such as model fit index and root mean square error of approximation, which are not comparable with the statistical coefficients generated by the remainder of the studies looking at the same association. Therefore, this study was excluded from the current meta-analysis. Out of the studies included, three of them (Deshpande et al., 2009; Veronese et al., 2016; Hsu et al., 2017) dichotomised the cognitive decline outcome, which was expressed as absence/presence of cognitive decline (where cognitive decline was defined as a decline of 3 or more points in the MMSE during the follow-up period), therefore their results provided a categorical index of association, expressed as Odds Ratios and pooled into a separate meta-analysis. The remainder of the studies looked at the MMSE as a continuous variable and expressed results as β coefficients and Standard errors, thus providing a dose-response association. Auyeung et al. (2011) divided their results by gender, therefore their results were considered as two different studies for the purpose of this meta-analysis.

The pooled ORs from Deshpande et al. (2009), Veronese et al. (2016), Hsu et al. (2017) showed that poor performance in walking speed at baseline was significantly associated with a greater risk of cognitive decline at follow up, OR= 2.26, 95% CI: 1.50–3.42, $p < .001$. There was no significant heterogeneity between studies ($I^2 = 0\%$, $p = .51$).

Study or Subgroup	log[Odds Ratio]	SE	Weight	Odds Ratio IV, Random, 95% CI
Deshpande 2009	0.8372	0.355	35.2%	2.31 [1.15, 4.63]
Hsu 2017	1.5217	0.6749	9.7%	4.58 [1.22, 17.19]
Veronese 2016	0.678	0.2836	55.1%	1.97 [1.13, 3.43]
Total (95% CI)			100.0%	2.26 [1.50, 3.42]

Heterogeneity: Tau² = 0.00; Chi² = 1.33, df = 2 (P = 0.51); I² = 0%
Test for overall effect: Z = 3.88 (P = 0.0001)

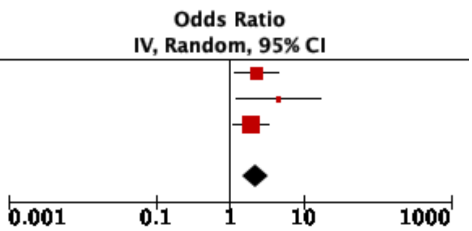

Figure 2: Forest plot for the meta-analysis of walking speed and dichotomised cognitive decline

The pooled mean estimate from the remaining studies looking at the association between walking speed and MMSE as a continuous variable showed a significant association between lowest performing groups in walking speed at baseline and cognitive scores at follow-up, β = .23, % CI = .03; .44, p =.030. There was significant heterogeneity between studies ($I^2 = 93\%$, p < .001). Sensitivity analyses were conducted to further explore the significant heterogeneity. The influence of single studies on the pooled association was investigated, by removing each study from the meta-analysis. It was found that no study, if removed from the analysis, substantially influenced the heterogeneity estimate. A symmetric funnel plot for the analysis (See Figure 4) suggests that publication bias was not present in the analysis.

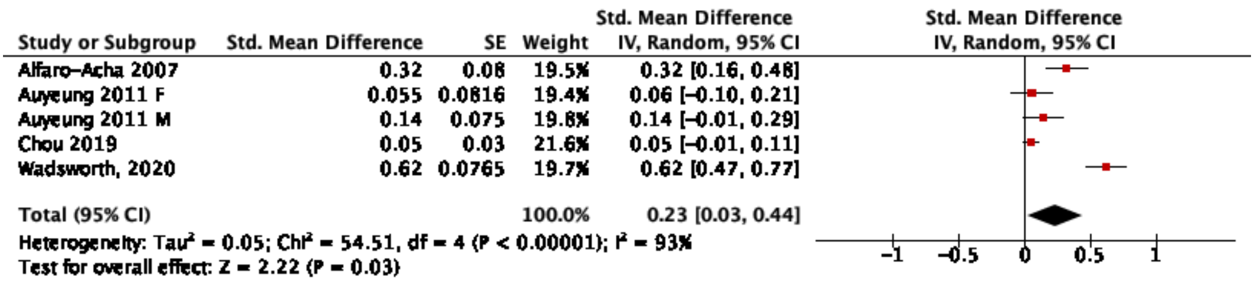

Figure 3: Forest plot for the meta-analysis of walking speed and continuous cognitive decline

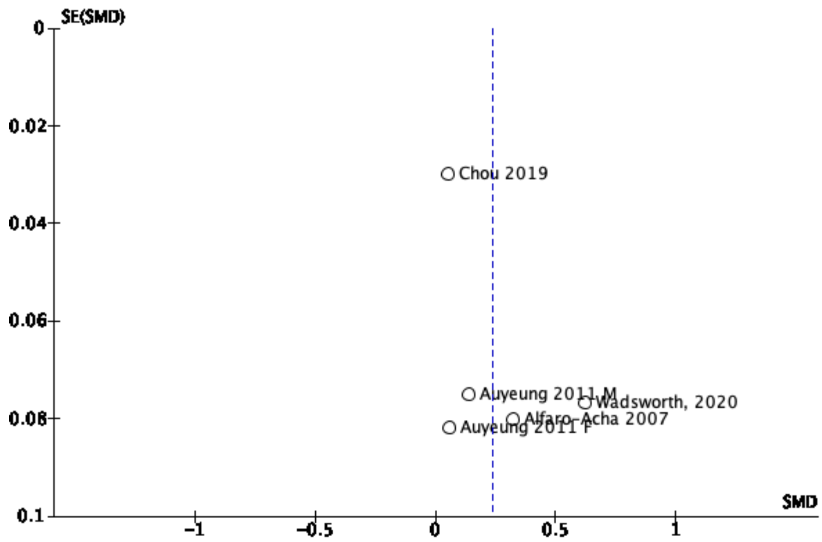

Figure 4: Funnel plot for the analysis between walking speed and continuous MMSE decline

Walking speed and the Digit Symbol Substitution Test (DSST)

Four studies were retrieved looking at the association between walking speed at baseline and DSST scores, which were assessed as a continuous variable at follow-up. DSST is a cognitive test which requires participants to match symbols to numbers following a set legend. Inzitari and colleagues (2007) express their results as ORs, therefore they were not grouped with the other 3 studies investigating this association. The remaining 3 studies (Best et al., 2016; MacDonald et al., 2017; Chou et al., 2019) expressed the results as β coefficients and SE, which were used to calculate a pooled estimate for this association. The analysis generated a significant model, β = .16, 95% CI = .01- .30, p = 0.04, which was characterised by very high and significant heterogeneity, I^2 = 91%, p < .001. Sensitivity analyses showed that this heterogeneity was driven by MacDonald and colleagues' (2017) study, which was underpowered (N= 121), and which followed participants for a significantly longer period of time than the other studies (25 years). The choice was made to remove this study from the analysis in order to reduce heteorgeneity, which yielded a significant model showing that poor performance in walking speed at baseline significantly predicted cognitive scores in the DSST, β = .20, 95% CI = .13- .28, p < .001. with no evidence of heterogeneity, I^2 = 0%, p = .360.

Study or Subgroup	Std. Mean Difference	SE	Weight	Std. Mean Difference IV, Random, 95% CI
Best 2016	0.19	0.04	83.5%	0.19 [0.11, 0.27]
Chou 2019	0.28	0.09	16.5%	0.28 [0.10, 0.46]
Total (95% CI)			100.0%	0.20 [0.13, 0.28]

Heterogeneity: Tau² = 0.00; Chi² = 0.84, df = 1 (P = 0.36); I² = 0%
Test for overall effect: Z = 5.60 (P < 0.00001)

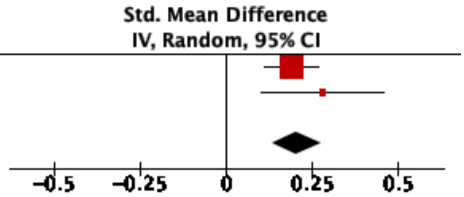

Figure 5: Forest plot for the meta-analysis of walking speed and continuous cognitive decline

Eight studies evaluated the association between handgrip strength and cognitive decline as assessed through the MMSE. As with walking speed, Sintjines and colleagues (2017) employed crossed-lag panel model estimates which are not comparable with the statistical coefficients generated by the other studies looking at the association between handgrip and MMSE, therefore this study was excluded from the current meta-analysis. Three studies (Veronese et al., 2016; Hsu et al., 2017; Kim et al., 2019) dichotomised the outcome, with the cut-off of a decline of 3 or more points in MMSE scores defining cognitive decline, or the absence thereof. These studies consequently expressed their results as Odds Ratios, and were thus included in a separate meta-analysis. Kim et al.'s (2019) study was excluded from this meta-analysis, as this study reported their results using the best performing quartile as a reference for performance, as opposed to Veronese et al. (2016) and Hsu et al. (2017), which used the lowest performing quartile as a reference for performance. The remaining studies looking at the association between handgrip strength and MMSE used β coefficients and Standard errors which were pooled into a different meta-analysis, that will be discussed below.

The pooled ORs from Veronese et al. (2016) and Hsu et al. (2017) showed that poorer handgrip strength at baseline was not significantly associated with greater risk of cognitive decline at follow up, OR= 1.30, 95% CI: .84–2.01, p = .240. Heterogeneity between studies was not detected, $I^2 = 0\%$, p =.90.

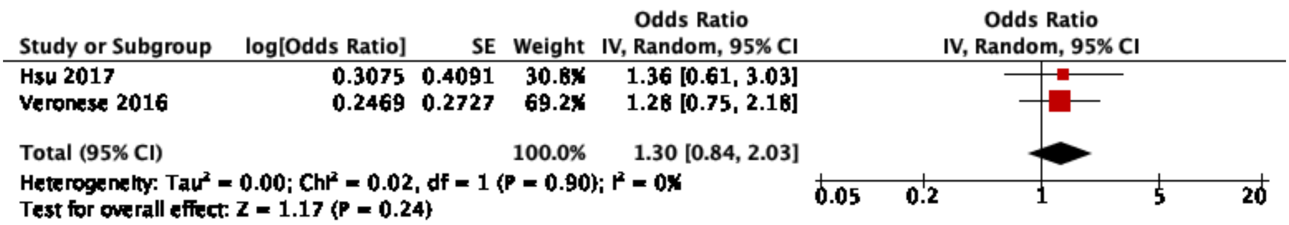

Figure 6: Forest plot for the meta-analysis of handgrip strength and dichotomised cognitive decline

Kim et al. (2019b) was excluded from the following meta-analysis because although they investigated cognition as a continuous outcome, they did not express their results using the worst-performing group as a reference for performance. The pooled mean estimate from the remaining studies looking at the association between baseline handgrip strength and MMSE as a continuous outcome showed that poorer handgrip strength at baseline was significantly associated with cognitive scores at follow-up, β = .18, 95% CI = .06; 0.29, p =.003. Heterogeneity between studies existed, I^2 = 73%, p = .010. Sensitivity analyses removing each study from the analysis revealed that the removal of Chou and colleagues' (2019) study did not have a significant impact on the overall significance, but it greatly reduced the heterogeneity, I^2 = 0%, p =.84. The funnel plot for the analysis (See Figure 8) shows a fairly symmetric distribution of the studies, suggesting the absence of publication bias.

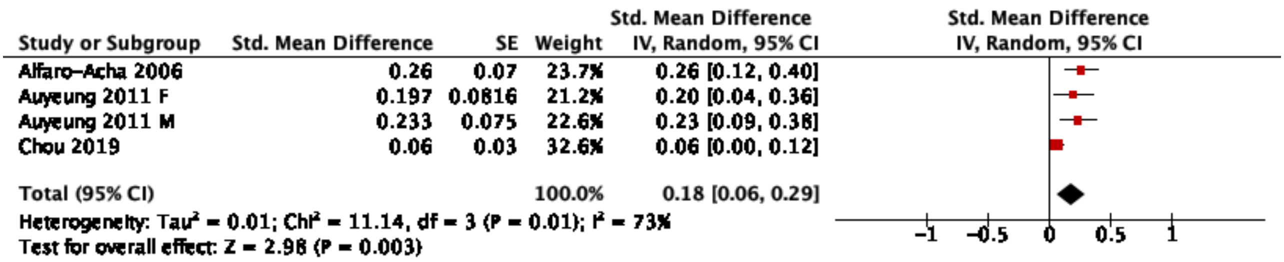

Figure 7: Forest plot for the meta-analysis of handgrip strength and continuous cognitive decline

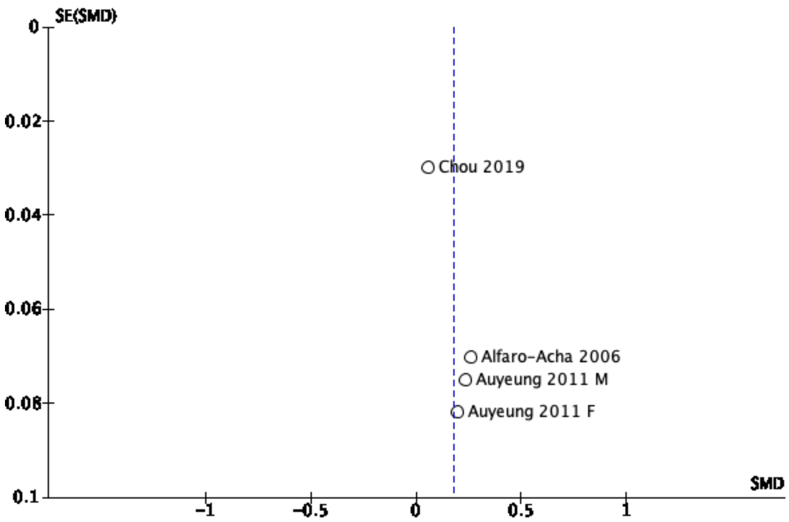

Figure 8: Funnel plot for the analysis between handgrip strength and continuous cognitive decline

Accelerometer-measured physical activity and cogniton as assessed through the Ascertain-Dementia 8-item questionnaire (AD8)

Three studies (Ku et al., 2017; Stubbs et al., 2017; Chen et al., 2020) were retrieved looking at the associations between physical activity as objectively assessed through accelerometer wearing, and cognitive scores as measured through the AD8, which rates memory, problem-solving abilities, orientation, and daily activities with a yes/no answer (higher scores represent greater cognitive decline). These studies expressed the outcome as Relative Risks, and Confidence Intervals. The other 3 studies assessing physical activity using accelerometer-based measures had heterogeneous and non-comparable outcomes, which made them inadequate for the purpose of being included in this meta-analysis. Additionally, out of the three studies expressing results as rate ratios, Ku et al. (2017) expressed their results as a function of sedentary behaviour, rather than daily physical activity, therefore their results were not comparable with the other two studies. The pooled RRs of the two studies looking at the association between accelerometer-based daily physical activity and cognitive decline showed that higher levels of daily physical activity as assessed through accelerometer were significantly associated with reduced risk of developing cognitive decline, RR= .70, 95% CI: .55–.88, p =.003. Heterogeneity was not detected, I^2 = 18%, p =.270.

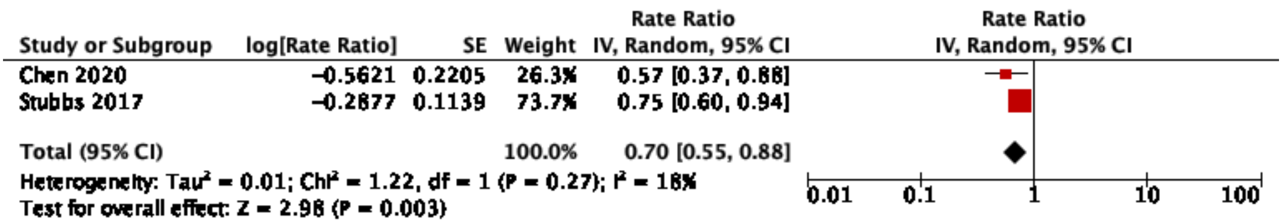

Figure 9: Forest plot for the meta-analysis of accelerometer-assessed daily physical activity and AD8

Balance and Mini Mental State Examination/Modified Mini-Mental State (3MS)

Two studies investigated the association between balance tests and MMSE (Wadsworth et al., 2020) or 3MS (Meunier et al., 2021). 3MS is a modified version of MMSE, which includes four additional items and a more specific level of scoring, and is designed to assess a greater scope of cognitive functions. MMSE and 3MS scores have been shown to be comparable (Teng & Chui, 1987), therefore outcomes from these measures were pooled in the same meta-analysis.

The results from this meta-analysis showed a significant association between balance and cognitive decline, whereby failure to perform in balance tests was significantly associated with decline in cognitive scores at follow-up, β = -.61, 95 % CI = -.81; -.41, $p < .001$. Heterogeneity was not detected, $I^2 = 62\%$, $p = .11$.

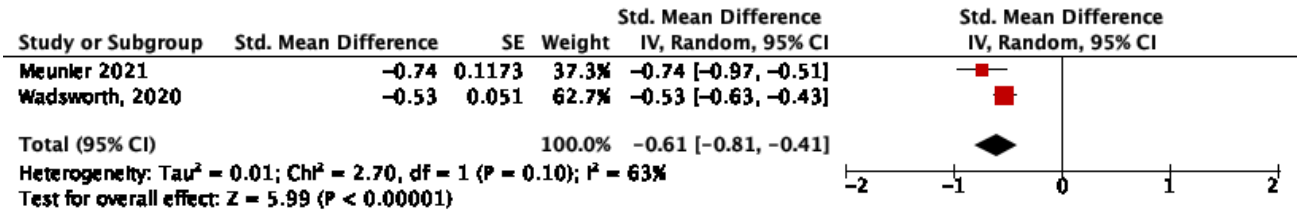

Figure 10: Forest plot for the meta-analysis of balance and continuous cognitive decline

Three studies investigated the association between sit-to stand test, which measures the time employed to complete a set number of chair stands, and MMSE scores. Auyeung et al. (2011) presented their results by gender, therefore the results were separated by gender and considered as two different studies for the purpose of this meta-analysis. Veronese and colleagues (2016) dichotomised the cognitive decline outcome, making their results not comparable with the other studies. Therefore, this meta-analysis only included results from Auyeung et al. (2011) and Wadsworth et al. (2020).

The results of the meta-analysis showed that low performance in the chair stand test was not significantly associated with cognitive decline as assessed through MMSE/3MS, β = -.33, 95% CI = -.70; .05 .19, p = .090. High heterogeneity existed in the analysis, I^2 = 95%, p < .001, which was significantly reduced when removing Wadsworth et al. (2020) from the analysis, I^2 = 62%, p = .100. This may be attributable to the significantly different lengths of the follow-up periods (4 years for both Auyeung and colleagues, 2011, and 9 years for Wadsworth and colleagues, 2020).

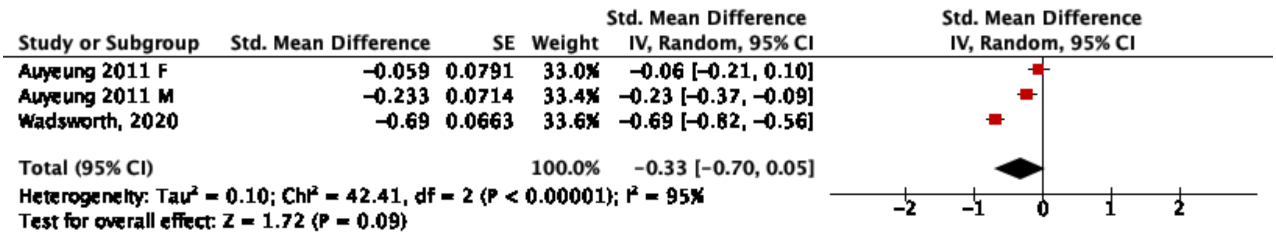

Figure 11: Forest plot for the meta-analysis of chair-stands and continuous cognitive decline

Narrative synthesis of the results

This section will discuss the results from our meta-analysis, considering our findings in light of the broader scope of the literature that was retrieved from our search. As mentioned thus far, some of the studies retrieved did not lend themselves for the purpose of being included in the above meta-analyses, due to the non-comparability of the outcome measures employed, or the heterogeneity of the statistical coefficients reported by studies using substantially different statistical techniques. Therefore, we will provide a narrative overview of the results from these studies, highlighting emerging issues or salient aspects in the existing body of research.

Walking speed and cognition

The results from our meta-analysis showed that higher performance in walking speed at baseline significantly predicted better global cognitive function at follow-up, as measured by results from the MMSE both as a dichotomised outcome, OR= 2.26, 95% CI: 1.50–3.42, $p < .001$, or as a continuous outcome, $\beta = .23$, % CI = .03; .44, $p =.030$, and by results from the DSST, $\beta = .20$, 95% CI = .13- .28, $p < .001$.These results support previous evidence showing that slow walking pace is significantly correlated with higher risk of cognitive deterioration and dementia onset in older adults (Verghese et al., 2007; Quan et al., 2017). Our literature search retrieved a total of 20 studies looking at the association between walking speed performance and cognitive decline at follow-up. Out of these studies, 11 were included in the meta-analyses as they could be categorised into different groups according to common outcomes measures (MMSE and DSST) and statistical indexes of association. The results from the remaining 9 studies that were not included in the meta-analysis appear to be mostly in line with our results. These studies showed that better walking speed performance at baseline was correlated with better cognitive outcomes in attention (Inzitari et al., 2007; Abe et al., 2017), memory (Mielke et al. 2013; Gale et al., 2014; Abe et al., 2017; Stijntjes et al., 2017; Okely et al., 2020), visuospatial function, verbal fluency (Abe et al., 2017), processing speed (Inzitari et al., 2007; Gale et al., 2014; Stijntjes et al., 2017), executive function (Mielke et al., 2013; Gale et al., 2014; Stijntjes et al., 2017), orientation (Ho et al., 2001), and global cognition as assessed through the Montreal Cognitive Assessment (MoCA) (Suzuki et al., 2015) at follow up.

Interestingly, Best and colleagues (2016) and Okely et al. (2020) assessed walking speed at different time points and found that early *decline in walking speed* predicted later decline in cognitive functioning. The approach of investigating whether changes in the physical domain predict changes in the cognitive domain is rarely employed in longitudinal studies, as it requires the assessment of physical capability at different time points. The investigation of physical performance as a changing variable, rather than the use of its baseline measure, was indeed sparsely used in the studies retrieved from our search (Best et al., 2016; Ritchie et al., 2016; Okely et al., 2020). Nonetheless, the strategy of investigating changing trends in physical performance to predict changes in cognitive performance is gaining growing popularity and should be further encouraged, as it provides valuable insight on the directionality of the relationship between physical and cognitive functions. Further, this approach provides a starting point to investigate the common resources that physical and cognitive processes share, which may underlie the parallel declining trend. Other studies such as Ritchie and colleagues (2016) and Zaninotto et al. (2018) observed inconsistent associations between walking speed and cognition, whereby the former study found that these functions exhibited similar changing slopes but one did not predict the other, and the latter found that baseline gait speed did not predict subsequent decline in cognitive functioning, questioning the common cause hypothesis of physical and cognitive decline in older age. While Ritchie and colleagues propose that their sample size may have lacked the appropriate power, and their study may have not had an adequate follow-up length to detect an association, Zaninotto et al. argue that the use of time-invariant variables of physical functioning may have concealed the relationship between physical functioning and cognitive decline. Nonetheless, the majority of the evidence retrieved from this search seemed to point towards an association between baseline walking speed and cognitive performance at follow-up, even in studies with similar follow-up lengths and sample sizes.

Walking speed tests are a widely used, objective measure of physical performance. The vast majority of studies measured walking speed over a distance ranging from 2 to 8 metres, and subsequently computed walking speed as distance/time (m/s). Some of the studies employed alternative methods of measuring walking speed, which allowed to broaden the scope of the investigation to other aspects of gait such as gait velocity standardised by average leg length, and stride time variability, which indicates the time elapsed between two footfalls of the same foot (MacDonald et al., 2017). Namely, MacDonald et al. (2017)) utilised the GAITRrite system, a computerised walkway deriving information on gait stride and variability through pressure-activated sensors (MacDonald et al., 2017). Interestingly, MacDonald and colleagues found that whilst simple walking speed was only a marginal predictor of cognition, GAITRite-assessed markers such as gait velocity and variability were significant, strong predictors of cognitive changes at follow-up. This evidence suggests that computerised walkway systems such as GAITRite may be particularly sensitive for measuring subtle age-related changes in walking speed, which may be measurable through highly specific markers of gait variability that simple timed walk tests may be unable to capture. The use of such devices in ageing research focusing on physical performance, and more specifically on the investigation on gait, should thus be encouraged.

Handgrip strength and cognition

Our meta-analysis of the studies which assessed handgrip strength as a continuous measure showed that poorer handgrip strength significantly predicts greater cognitive scores at follow-up, β = .18, 95% CI = .06; 0.29, p =.003. On the other hand, the results from the meta-analysis considering cognitive decline as a dichotomous outcome showed no significant association between handgrip strength at baseline and cognitive status at follow-up, OR= 1.30, 95% CI: .84–2.01, p = .240. It must be noted that the results from the latter meta-analysis must be interpreted with caution, as they consisted of only two studies which shared comparable cognitive measures and similar statistical coefficients of association, as opposed to a more extensive range of evidence from five studies included in the former analysis, which did indicate the presence of an association. The results of the meta-analysis were likely driven by the non-significant effects from Hsu et al.'s (2017) investigation, which did not find handgrip strength to be a significant risk factor for cognitive decline. Importantly, this study followed up 249 male participants during one year, therefore the lack of a significant effect may be justified by the reduced variability in handgrip scores driven by the sample being entirely male, or the brief duration of the follow-up period, which may not have been sufficient for associations to emerge.

Our literature search found 16 studies using handgrip strength as a predictor of cognitive performance. The included studies consistently assessed handgrip strength using hand-held dynamometers. Of these studies, 7 were included in the meta-analyses discussed because they all used the MMSE to assess cognitive performance, while the remaining 9 could not be included as they did not share comparable outcome measures or statistical techniques. Boyle et al. (2009) and Buchman and colleagues (2012) used handgrip strength to compute a muscle strength score and a composite frailty measure respectively, without analysing the predictive role of handgrip strength in isolation, therefore we cannot consider their results for this specific measure. The remaining studies showed that handgrip strength at baseline was significantly associated with changes in global cognitive function (Stijntjes et al., 2017; Kim et al., 2019), memory and processing speed (Stijntjes et al., 2017), executive and visuospatial function (Viscogliosi et al., 2017), and verbal memory (Okely et al., 2020).

The remainder of the studies that we retrieved found inconsistent associations between grip strength and cognition. For instance, Abe and colleagues (2017) did not find handgrip strength to predict cognitive decline at follow-up. Instead, their results showed that a peg-moving task assessing hand dexterity was strongly predictive of cognitive decline at follow-up. The authors thus hypothesised that a measure of fine hand motor function, as opposed to grip strength, may be able to detect subtle changes in cognition, and hold a predictive power that may be more sensitive to cognitive changes in older age. Whilst this hypothesis could be a valid explanation for these results, there are no other studies in our search which corroborate these findings, nor which use fine motor dexterity as a predictor of cognitive impairment. However, later evidence has since supported strong associations between fine hand dexterity and cognition (Kobayashi-Cuya et al., 2018), in light of the greater involvement of cognitive processes such as planning and attention, that are required for precise hand dexterity as opposed to grip strength. Given the vulnerability of hand dexterity to ageing, as well as its potential to predict subtle changes in cognition, this measure of physical function should be further implemented in future ageing research. Further, its efficacy in predicting cognition should be analysed against that of handgrip strength, in order to substantiate claims that hand dexterity might be a more accurate measure to employ in ageing research. In addition to these results, both Ritchie and colleagues (2016) and Suzuki et al. (2017) did not find handgrip strength to significantly predict cognitive decline in later life. In both cases, the authors propose that this may be due to the inadequate sample size or follow-up length of their study. However, as pointed out by Suzuki and colleagues in their discussion, even a follow up length of one year, as was employed in their own investigation, is considered to be a relevant follow-up period from a public health research perspective. Here, we propose that the variety of the cognitive outcomes employed by these studies may underlie the inconsistency in the results, whereby handgrip strength seems to predict changes in cognition more consistently when using the MMSE as a cognitive outcome.

Interestingly, an aspect of ageing research that is gaining increasing popularity is the debate focusing on the directionality of the association between grip strength and cognition. Establishing the directionality of this effect carries great significance for its potential implications in clinical practice, as it would provide understanding of the mechanisms underlying ageing and inform the development of interventions and diagnostic tools. However, research looking at the temporal relationship between these variables has yielded inconsistent results thus far: while the studies discussed above point towards a predictive role of handgrip strength on cognition, there is evidence in the literature showing that suboptimal cognition at baseline might be predictive of poorer cognitive impairment at follow-up (Taekema et al., 2012). In the studies included hereby, both Stijntjes and colleagues (2017) and Kim et al. (2019) examined the temporal relationship between handgrip strength and cognition in older adults in both directions. While the former study showed that the temporal association between cognition and grip strength differs across domains and age groups, the latter investigation found the association to be bi-directional only in patients with cognitive impairment at follow-up. From the evidence that we gathered, the temporal relationship between handgrip strength and cognition in older adults appears to be inconclusive, although only two studies assessed this association in both ways, in spite of the clinical significance of establishing its direction. Nonetheless, the finding of a bi-directional relationship between these two variables in adults with cognitive impairment is promising, as it suggests that interventions targeting either cognition or physical functioning may have mutual beneficial effects on the other variable (Kim et al., 2019). Additional research looking at bi-directional associations will further elucidate on this association.

Accelerometer-assessed physical activity and cognitive decline

The results from our meta-analysis showed that higher levels of daily physical activity as assessed through accelerometer were significantly associated with reduced risk of cognitive impairment, RR= RR= .70, 95% CI: .55–.88, p =.003. This meta-analysis was composed of two studies with comparable outcomes and shared statistical indexes of association, out of the total of six studies that were retrieved using accelerometer-based measures of physical function. The results that were not included in the meta-analysis found total daily physical activity to be associated with the rate of global cognitive impairment (Buchman et al., 2012; Ku et al., 2017; Zhu et al., 2017), as well as with episodic memory (Buchman et al., 2012; Halloway et al., 2017; Zhu et al., 2017), semantic memory (Buchman et al., 2012; Halloway et al., 2017), executive functioning (Zhu et al., 2017), and perceptual speed (Buchman et al., 2012).

Accelerometer-based measures of physical activity provide an objective measure of daily activity, which is highly tolerated by older adults. In light of these characteristics, accelerometers are a promising objective measure of physical functioning that is gaining increasing popularity in epidemiological studies. However, there are some problematic aspects to consider when implementing the use of these devices in ageing research, which emerged from a closer inspection of the studies retrieved from our search. Firstly, accelerometer-based measures require a high degree of participant adherence, as they are typically assessed over a minimum period of 7 days, and the device needs to be constantly worn in order to provide an accurate measure of daily physical activity (Gemmill et al., 2011). This was reflected by the relatively low sample sizes observed in the studies described above, which included sample sizes as low as 59 participants, and high drop-out rates, with Zhu et al. (2017) reporting drop-out rates of nearly 50% at follow-up. Exclusions are also common due to failure to return or wear the accelerometer, device or user errors, or noncompliance with the required wear time criteria (Zhu et al., 2017). These aspects could be partially mitigated by developing protocols to ensure that participants are cognisant on how to use the device.

Secondly, it is important to note that the studies discussed here differed in the way that they assessed total daily activity. While most of the studies discussed here assessed total daily physical activity, one study used the device to compute an index of sedentary behaviour (Ku et al., 2017). Additionally, among the studies using accelerometer data to compute a measure of daily activity, studies adopted diverse thresholds to categorise participants into different activity groups (low, moderate, high) based on their daily step count, which makes cross-study comparisons problematic as activity levels tend to overlap. Although the results from the studies discussed here seem to indicate that accelerometer-based measures may be a good predictor for later cognitive performance, the issues discussed thus far need to be addressed in future research. Standardised outcome measures, methods to reduce drop-out rates, and thresholds for the categorisation of physical activity levels need to be established by future epidemiological research, in order to obtain comparable datasets that allow to further substantiate the predictive role of accelerometer-based measures of physical functioning to assess changes in cognition.

Balance and cognitive decline

According to the results from our meta-analysis, baseline balance scores were associated with greater cognitive decline at follow-up as measured through MMSE scores, $\beta = -.61\%$ CI = -.81; -.41, $p < .001$. The remaining study using balance as a predictive measure, which was not included in the meta-analysis due to the use of a non-comparable cognitive outcome, showed that balance did not significantly predict cognitive decline at follow up (Abe et al., 2017). While the implications of these results are not discussed by the authors, we hereby hypothesise that the failure to observe a significant association may be due to inconsistencies in the measures used to assess balance. Indeed, although different methods to evaluate balance are universally referred to as "balance" in epidemiological research, some of the tasks used to assess this physical function may be more challenging than others. Whilst some measures involve participants maintaining a full tandem stance (i.e. standing with one foot in front of the other), a semi-tandem stance (i.e. one foot placed slightly in front of the other) or a side-by-side stance for as long as possible, other studies evaluate one-leg standing balance, (i.e. the amount of time that participants are able to stand on one leg) as an index of balance. Although both measures provide an indication of the same physical function, the latter is a more demanding task, which requires greater physical performance, that may not be granted at an older age. In the studies that we retrieved, only Abe and colleagues used the latter assessment of physical function, which may explain why their study did not show a significant association between balance and cognitive performance. It is thus relevant to consider the implications of comparing different measures of the same construct, as well as to select a suitable measure for ageing research.

This section will also discuss results from the timed up and go (TUG) test, which is considered to be a measure of dynamic balance and overall mobility (Abe et al., 2017). The current review did not include a meta-analysis for this test, since only three studies were retrieved using TUG as a predictor of cognitive performance, which all employed different outcome measures. Of these studies, Abe et al. (2017) found that baseline TUG performance was significantly associated with global cognition as measured through a composite score obtained from 5 cognitive tests, Katsumata and colleagues (2011) showed TUG performance at baseline to predict global cognitive decline as assessed through the J- MMSE, and Suzuki et al. (2015) found that TUG predicted global cognition as measured through the MoCA-J. Although there is still relatively little evidence on TUG as a predictive factor for cognitive decline in older age, associations between baseline TUG scores and subsequent executive decline have been previously shown in adults with MCI (McGough et al., 2011), and have been further supported by the studies on healthy participants that we retrieved.

Chair stands and cognitive performance

The results of our meta-analysis showed that poor performance in the chair stand test was not significantly associated with cognitive decline as assessed through MMSE/3MS, $\beta = -.33$, 95% CI = -.70; .05 .19, $p = .090$. Findings from the other studies retrieved from our search which used chair stands as a predictive measure of cognitive performance seem to corroborate these results. Both Veronese et al. (2016) and Abe and colleagues (2017) failed to show a significant association between chair stands at baseline and subsequent cognitive decline. The chair stand test is not attributable to a specific physical function: it encompasses leg strength, endurance, and components of balance. As such, performance in this test provides a composite measure of different aspects of physical performance, which makes it difficult to interpret these results. It might be worth considering that when more physical functions are acting in concert for the execution of a complex physical task, predictive relations may be harder to discern. Nonetheless, the evidence gathered from our search seems to point against a predictive role of chair stands test for cognitive decline.

Cardiorespiratory measures and cognitive performance

Our search retrieved three studies using cardiorespiratory measures at baseline to predict cognitive performance at follow-up in older adults (Barnes et al., 2003; Ritchie et al., 2016; Okely et al., 2020). These studies yielded somehow inconsistent results, which were not pooled into a meta-analysis as the studies all used different cognitive outcomes. In terms of assessing cardiorespiratory function, the retrieved studies most commonly employed Forced Expiratory Volume (FEV), a measure of lung function which refers to the volume of air exhaled in the first second during forced exhalation after maximal inspiration. Both of the studies using this method to assess cardiorespiratory fitness failed to observe a significant correlation between FEV and cognitive outcomes at follow-up. More specifically, Ritchie et al. (2016) found no significant evidence that baseline FEV could predict cognitive changes in older adults, whereas Okely and colleagues (2020) did not find any significant correlation between changes in FEV and changes in cognition in older adults.

A more comprehensive investigation of cardiorespiratory function was performed by Barnes and colleagues (2003), who used treadmill exercise tests to investigate peak oxygen consumption (peak VO_2), treadmill exercise duration, and oxygen uptake efficiency slope (OUES), defined as the relation between oxygen consumption and minute ventilation during exercise. Results from this investigation showed that participants with poorer cardiorespiratory fitness at baseline had greater cognitive decline in global cognition and in measures of attention and executive function at follow-up. When considering these results globally, it would appear that a more extensive assessment of cardiorespiratory function may have a higher testing resolution, allowing to predict cognitive changes more accurately, while static FEV measures may deflate the association between cardiorespiratory fitness and cognitive decline. Further, while FEV provides a measure of static lung function, VO_2 is a performance-focused measure, which informs on oxygen utilisation during exercise. Therefore, VO_2 might provide a more accurate picture of physical performance, which may explain the inconsistency of the results discussed above.

Composite measures of physical function

Of the included studies, Veronese et al. (2016) also assessed physical function using the Short Physical Performance Battery (SPPB), a composite measure derived from three tests assessing the ability to complete tasks on balance, gait speed and chair rises. While this score was not pooled into any of the meta-analyses in light of it being a composite measure, SPPB-based results from this study supported an modest predictive role of performance-based physical function on cognition, with lower scores in the SPPB predicting worse cognitive outcomes. However, this association was no longer significant after the adjustment of the analysis for potential covariates, suggesting that the SPPB may not be a strong predictor of cognitive decline. The involvement of a chair-stands component in the SPPB could potentially explain why the SPPB was a weak predictor of cognition in Veronese and colleagues' (2016) investigation. In fact, our analysis did not find chair rises to be a significant predictor of cognitive decline. We attributed these findings to this measure assessing the synchronous functioning of a variety of functions such as leg strength, endurance, and components of balance, which wakes predictive relations harder to discern due to the diversity of aspects of movement involved. Therefore, the inclusion of this task in the SPPB may be driving the only modest association with cognitive decline that we observed in our analysis.

Discussion

The objective of this review was to evaluate whether different performance-based measures of physical function at baseline had any predictive power on cognitive decline at follow-up in adults aged 50 or older. The results from our meta-analyses suggested a longitudinal association between measures of walking speed, handgrip strength, accelerometer-based physical activity, balance, timed up and go test and global cognition. These associations appeared to be confirmed when considering our findings in the broader context of the literature that was not included in the meta-analyses due to non-comparable outcomes. Indeed, the studies retrieved supported further associations between the abovementioned physical function measures and memory, executive functioning, processing speed and visuospatial functions. Nonetheless, the variety of outcome measures used made cross-study comparisons problematic. The results for cardiorespiratory measures and the chair-stands test yielded inconsistent associations, which further indicate the need for more extensive, standardised testing procedures with higher testing resolution. In this section, the possible mechanisms that may mediate the association between physical functioning and cognitive performance will be reviewed, along with considerations on common themes on methodological issues in the testing procedures that emerged from our search. Finally, suggestions for future research will be outlined to address these problematic aspects, and to expand on promising, innovative devices and analysis methods to objectively assess the physical functions that we identified in our search.

Potential mechanisms mediating the association

There are different potential mechanisms underlying the association between the physical functions that we examined and cognition. Inflammatory processes are known to be significantly related to cognitive impairment (Cesari et al., 2004; Glade, 2010). Skeletal muscles, which connect our bones and allow us to perform a variety of movement and functions, are known to have a role in the secretion of inflammatory mediators such as cytokines and interleukin-6, which are involved in inflammatory processes, as well as in the loss of muscle strength (Pedersen & Febbraio, 2012; Cui et al., 2021). Skeletal muscle also plays a role in the secretion of brain-derived neurotropic factor, a protein which regulates neuronal growth and neural plasticity. Therefore, the loss of muscle mass mediated by age-related physiological changes and by the release of inflammatory mediators may reduce the secretion of BDNF, and may potentially mediate the relation between the physical functions that we investigated and cognitive decline.

Associations between walking speed and cognition have been established by previous reviews showing that slow walking speed is predictive of both cognitive decline (Quan et al., 2017) and greater dementia risk (Cooper et al., 2011; Kikkert et al., 2016; Quan et al., 2017). Walking speed is known to be significantly associated with muscle loss, which in turn is strongly correlated with inflammatory processes which, as discussed above, are related to cognitive dysfunction (Cesari et al., 2004; Quan et al., 2017). Further, walking is a demanding activity which requires the coordination of complex motor and cognitive processes involving the optimal working of cerebral and cerebellar structures (Callisaya et al., 2017). Therefore, walking speed may be in and by itself, an indicator of concurrent cognitive status, which may consequently explain later cognitive dysfunction (Quan et al., 2017). Further, slow walking speed may be an indicator of overall physical inactivity, which is known to be associated with worse cognitive outcomes (Quan et al., 2017).

Similar mechanisms may explain the correlation between accelerometer-based measures of daily steps and cognition, which was supported by the findings from our meta-analysis. These findings were in line with previous research showing cross-sectional associations between daily steps and cognition (Calamia et al., 2018); however, there is still a paucity of longitudinal evidence looking at accelerometer-based measures of physical function and cognition. It is widely believed that this association may be due to the impact that aerobic exercise may have on brain function. Daily step count is known to significantly contribute to aerobic energy expenditure (Tudor-Locke et al., 2011). In turn, aerobic exercise has been shown to facilitate the release of BDNF (Firth et al., 2018), as well as promoting improvements in brain structure and brain matter volume (Colcombe et al., 2006). Therefore, individuals engaging in a high daily step count as assessed through accelerometer measures may show better cognitive outcomes by means of the benefits of aerobic exercise on neural structure. The association between low handgrip strength and greater cognitive decline that we observed in our meta-analysis corroborates findings from previous reviews showing low handgrip strength to be predictive of both cognitive impairment and dementia (Cui et al., 2021). Handgrip strength is strongly correlated to muscle loss, therefore the inflammatory processes discussed above, along with the reduced release of BDNF, may mediate the association that we observed in our meta-analysis. Further to this, there is evidence indicating that white matter integrity may be correlated with handgrip strength (Sachdev et al., 2005). Therefore, poor grip strength may be an early marker of cognitive decline reflecting suboptimal white matter functioning.

Our meta-analysis showed an association between tandem balance measures, as well as the timed up and go test, which is considered to be a measure of dynamic balance and mobility, and cognition.

These findings are in line with previous research showing that poor performance in balance measures was associated with higher risk of developing dementia (Bullain et al., 2016), and that dynamic balance was associated with reduced risk of executive decline (McGough et al., 2011). However, the mechanisms underlying the association between balance and cognition are complex and not yet fully understood. It is believed that the vestibular system may mediate the association between standing balance and cognition (Meunier et al., 2021). The vestibular system regulates balance through the integration of proprioceptive and brain feedback (Massion, 1998). There is evidence showing that vestibular dysfunction is associated with poor balance and cognitive impairment (Smith et al., 2013), suggesting that this system may be the common cause underlying this association.

Common themes in methodology: issues of heterogeneity and consistency

Our meta-analyses did not show cardiorespiratory measures of physical function, nor chair stands test, to have a predictive role on cognition. As for cardiorespiratory measures, Ritchie et al. (2016) and Okely and colleagues (2020) found no significant associations between FEV and cognition in older adults. Conversely, Barnes and colleagues (2003) found peak VO_2 during treadmill exercise to be predictive of global cognition at follow-up. This would appear to suggest that a more sensitive, performance-focused measure of physical function such as peak VO_2 may offer a more accurate prediction of cognition, as opposed to a static measure such as FEV which may not be as sensitive and may be deflating the magnitude of such association. In terms of the chair-stand test, the meta-analysis only consisted of two different studies. This is owed to the studies with chair-stand test as a predictor employing outcome measures assessing different cognitive functions, which did not offer the possibility to compare the results, as well as to chair-stands tests being a complex measure, involving different aspects of physical function such as dynamic balance and overall mobility. A closer inspection of these results, when considered in light of additional issues emerging from the testing procedures of the studies that we included, are indicative of issues with the heterogeneity and the consistency of the methodology adopted by these studies. We will now focus on three problematic areas identified in this field of research: 1) heterogeneity in the outcome measures used, 2) low resolution and inconsistencies of the predictors and outcome measures, 3) lack of standardised guidelines for methodology.

Heterogeneity in the outcome measures

The issue of heterogeneity in testing procedures was recurrent in the sample of studies retrieved from our search, and was a contributing factor to the structure of our meta-analysis. While the majority of studies used the MMSE to investigate global cognition as a primary outcome measure, there was still a significant portion of studies looking at different cognitive functions ranging from executive function, to memory, visuospatial functions, processing speed, attention, and verbal fluency. Additionally, there was significant variety in the measures employed to assess the same cognitive construct. For instance, both Abe and colleagues (2017) and Okely et al. (2020) investigated visuospatial function. While the former study used the clock drawing test to assess visuospatial functioning, the latter investigation used the Matrix Reasoning and Block Design subtests from the from the Wechsler Adult Intelligence Scale to measure the same cognitive construct. Although this is an effective example to illustrate the heterogeneity, this issue was not limited to visuospatial function and was encountered consistently in the studies that we retrieved. This heterogeneity in outcomes made cross-study comparisons difficult, given that different neuropsychological tests may tap on diverse aspects of the same cognitive function. Additionally, neuropsychological tests may exhibit different sensitivities in detecting impairments in a specific function (Proust-Lima et al., 2007). This is especially problematic in the ageing population, where some tests may be more sensitive in participants with lower levels of cognition than others (Proust-Lima et al., 2007). Our meta-analysis aimed at reducing heterogeneity by grouping studies according to common outcome measures. While this allowed for greater consistency in the analyses, it also translated into a significant portion of the studies being excluded from the analyses and only being discussed narratively. Moreover, conducting multiple meta-analyses causes greater risk of introducing type I errors, caused by results reaching significance by chance, because the probability of finding at least one significant model due to chance increases in parallel with the number of models being tested. Nonetheless, the benefit of conducting multiple meta-analyses with reduced rates of heterogeneity, as well as consistent variables and statistical indexes of association was deemed to outweigh the increased risk for statistical errors introduced by multiple testing. Indeed, the approach of conducting multiple meta-analyses ensured that the studies could be meaningfully pooled together, in order to explore specific associations in greater depth. Future research may consider the use of multi-level meta-analysis to counter the issue of multiple testing. In multi-level meta-analysis, multiple levels of data structure are considered. In particular, meta-analytical data has an innate hierarchical structure whereby differences exist at the participant level, but participants are also nested within studies and differences at the study level are also considered. Multilevel meta-analyses may consider additional levels that may create dependency between effect sizes (i.e. effect sizes that are correlated), which may artificially reduce heterogeneity and lead to false significant results. Dependency of effect sizes may be due to studies using different methods to assess the same variable, or to studies being conducted in different countries. Multilevel meta-analysis allows to account for these differences by integrating a third level into the structure of a meta-analysis, in which studies are nested according to similar characteristics such as region where the study was conducted or method used to assess physical function. While this approach would certainly account for different sources of heterogeneity, the variety of the physical functions measured, the different outcomes measures employed, as well as other sources of heterogeneity such as follow-up length and region of the study would require multiple multi-level analyses with multiple levels, resulting in a complex approach. For this reason and in light of the existing heterogeneity, the approach of multiple testing adopted here seems reasonable.

In light of the existing heterogeneity, the adherence to homogeneous, standardised testing guidelines should be promoted, in order to facilitate cross-study comparisons and allow to conduct comprehensive meta-analyses, resulting in a thorough evaluation of cognitive functions.

Issues with consistency and resolution of predictive and outcome measures

Further to the heterogeneity issues discussed above, the exclusive reliance on the MMSE to capture cognitive changes in the elderly may be problematic. This test is the most commonly used screening tool to measure cognitive changes in older adults, and its wide use in longitudinal research allowed for cross-study comparisons in our meta-analyses. However, there is evidence indicating that small changes in MMSE scores may be due to measurement error or practice effects (Hensel et al., 2007). Given that a large number of the studies that we included considered small changes of as little as three points in the MMSE as the criterion indicating cognitive decline, the reliability of such small changes should be questioned. With other studies using as many as 13 or 19 cognitive tests (Okely et al., 2020 and Buchman et al., 2007; 2012 respectively) to compute global cognitive scores, the issue of heterogeneous resolution in testing procedures should be addressed by future longitudinal research. Indeed, a procedure relying on 19 cognitive tests evaluating different areas of cognition may provide a more accurate picture of the cognitive status of an individual, which may also hold greater sensitivity to detect cognitive changes. While a battery of 19 tests is auspicable, it may not always be a feasible method to introduce in longitudinal research due to the time cost of such extensive testing. However, there are alternative testing procedures adopted by the studies included in this review that might be more attainable. For instance, Abe and colleagues (2017) employed the “5 Cog” testing battery, which investigates a reasonable range of cognitive functions such as attention, memory, visuospatial function, verbal fluency and reasoning through 5 cognitive tests. The implementation of a more extensive testing protocol, that is homogeneous across studies, may lead to higher testing resolution and greater possibility to make cross-study comparisons in longitudinal research.

The use of innovative techniques to assess physical function highlighted how some methods to measure physical function may lack the required testing resolution to capture salient aspects of physical variables. This was exemplified by the use of innovative techniques to investigate gait. For instance, MacDonald et al. (2017) employed the GAITRrite system, a computerised walkway assessing different aspects of gait through a sensor-based system. As previously discussed, MacDonald and colleagues found that GAITRite-assessed markers of gait were strong, significant predictors of cognitive changes at follow-up, whilst simple walking speed measures only marginally predicted cognition. This would appear to suggest that computerised methods to assess gait may inform us on aspects of gait with strong predictive potential for cognition in older adults, whereas walking speed, which is considered a salient predictor of cognition, may still be a less sensitive measure when compared to more innovative methods. On the other hand, the use of accelerators in ageing research has also yielded promising results, and may be a more sensitive method to assess walking habits than walking speed tests. The results from a meta-analysis from two studies (Stubbs et al., 2017; Chen et al., 2020), along with the findings from three more studies that were narratively discussed (Buchman et al., 2012; Ku et al., 2017; Zhu et al., 2017), all point towards an association between accelerometer-based physical activity and cognition. These devices can be easily worn and provide an objective, reliable indication of daily physical activity, that is not dependent on test performance measures but objectively reflects daily walking habits. As knowledge on modifiable risk factors for cognitive impairment becomes increasingly relevant in light of the ageing population, the use of such devices should be encouraged in order to obtain accurate measurements of physical function, and make reliable projections on its impact on cognition. As previously highlighted, there seems to be an issue with cut-off points to define the separating line between physical activity levels (low, moderate, high), where these thresholds vary significantly across studies, leading to conflicting findings in the literature (Trost & O'Neil, 2014). This issue was described by Lyden and colleagues (2011), who reviewed the validity of 11 energy expenditure predictive equations for accelerators, and found that physical activity levels were misclassified 20-35% of the time using existing predictive models. An emerging resolution to this problematic aspect is a data reduction approach, which uses machine learning to extract patterns in accelerometer data, and utilises information on these patterns to compute intensity levels of activity (Trost & O'Neill, 2014). While still in its early stages, this approach may be a promising alternative to promote homogeneous guidelines for accelerometer-based measures of physical function.

Additionally, some issues relating to the consistency of the predictive measures emerged. In relation to balance, some studies used different variations of tandem stance to measure balance capability (Wadsworth et al., 2020; Meunier et al., 2021), while others looked at one-leg standing balance (Abe et al., 2017), which demands greater levels of physical performance. Studies using one-leg standing balance did not find balance to significantly predict cognitive functioning, while studies employing the former method did show this association. While other reasons may account for these results, the inconsistency in the predictive measures used may have an impact on the association between these functions. Similarly, studies assessing cardiorespiratory function through FEV (Ritchie et al., 2016; Okely et al., 2020), a measure of static lung function failed to show a predictive role of cardiorespiratory function on cognition, while studies using VO_2, a performance-focused measure of oxygen utilisation, did support this association (Barnes et al., 2003). Again, this may suggest that inconsistencies in the predictive measures used may conceal or inflate existing associations, and impact our understanding on the relation between physical functions and cognition.

Lack of standardised guidelines for methodology

The third and final issue that emerged from our search was the lack of standardised guidelines for methodology. As well as having heterogeneous testing procedures as discussed above, the studies that we included were characterised by different follow-up periods, used analysis methods yielding different indexes of association (either categorical or continuous), and included diverse covariates combinaions in their analyses. While a certain degree of methodological heterogeneity is to be expected on follow-up periods, the adherence to standardised guidelines indicating statistical methods and essential covariates that are to be included in the analyses should be encouraged. We attempted to mitigate for the lack of guidelines on covariate inclusion with our quality assessment checklist (see p. 37), by rating the inclusion of essential covariates in the analyses, which were established as age, gender, education levels and comorbidities. These essential criteria were met by the majority of the studies that we included here, and ensured that a standard for covariate inclusion was met. However, the results reported by each study included a variety of models with alternative covariates, the combinations of which were different across studies. This aspect becomes problematic when comparing results stemming from different models, which were obtained through the adjustment for diverse combinations of covariates. As for the statistical analyses, the majority of the studies included used regression models to look at the association between cognition and its predictors. Nonetheless, a number of studies (Ritchie et al., 2016; Stijntjes et al., 2017; Okely et al., 2020) used cross-lagged panel models, looking at reciprocal associations between physical function measures and cognition. As ageing research is evolving towards looking at reciprocal associations between predictors and outcomes in order to ascertain the direction of the observed associations, it is likely that this statistical method will become increasingly popular.

While this review is based on the hypothesis that baseline physical function levels may predict cognition, a growing amount of research is indeed focusing on associations between dynamic rates of change in these variables (Best et al., 2016; Ritchie et al., 2016; Stijntjes et al., 2017; Kim et al., 2019; Okely et al., 2020). The concept of associated changing slopes in both variables would lend support to the common cause hypothesis (Christensen et al., 2011), which holds that a common physiological ageing process may account for declines across physical and cognitive functions. This common process may be exhibited as declining trajectories across both physical and cognitive functions. While some of the studies that we retrieved found shared declines in physical functioning and cognitive trajectories (Best et al., 2016; Kim et al., 2019; Okely et al., 2020), other studies did not report a similar association (Ritchie et al., 2016; Stijntjes et al., 2017), suggesting that these dynamic trajectories may vary according to the physical and cognitive functions that are being investigated. Nonetheless, studies looking at associations between dynamic rates of change of physical and cognitive functions provide valuable information as to whether physical functioning and cognition change in concert over time, how strong the change-to-change association may be, and the directionality of said association. Although investigating change-to-change trajectories means that multiple measurements of physical function must be reported, this type of analysis is strongly encouraged in order to improve our understanding on the underlying mechanisms of this association.

This study had some strengths. Firstly, our analysis focused on longitudinal studies: when investigating outcomes that are related to the ageing process, it is essential to study these outcomes longitudinally, as proposed by Hofer and Piccinin (2009). Our review considered objective indicators of physical function rather than self-reported measures, in an attempt to reduce the recall bias that characterises the latter. Further, we considered a variety of physical tests, in order to gather extensive evidence on a multifaceted construct such as physical function, which is composed by different aspects, namely muscular strength, aerobic activity, and endurance. The physical function predictors that we looked at (walking speed, handgrip strength, balance, cardiorespiratory measures, sit-to-stand test, TUG test) effectively capture various facets of physical functioning, and therefore inform us on this variable comprehensively. As per the cognitive outcomes, we chose to gather evidence on cognitive decline, as opposed to focusing on outcomes based solely on dementia diagnosis (Hamer et al., 2009). Therefore, our approach considers preclinical stages of cognitive impairment, which allows to capture early stages of cognitive deterioration more accurately. Finally, our analysis attempted to reduce heterogeneity, by grouping studies together according to common predictors and outcomes- this allowed to investigate the association between each predictor and cognitive outcome in a reliable, thorough manner.

There are also some limitations to our study. While evidence from longitudinal studies allows to investigate patterns of association between physical function and cognition, it cannot inform on the causal relationship between these variables. Our analysis points to an association between baseline levels of physical tests and subsequent cognitive impairment; however, a causal relation cannot be inferred from evidence gathered from observational, longitudinal studies. There were some attempts to disentangle the direction of this association in the studies that we retrieved (Best et al., 2016; Ritchie et al., 2016; Stijntjes et al., 2017; Kim et al., 2019; Okely et al., 2020), with inconsistent results reported across different studies for the same physical and cognitive functions. Secondly, while the strategy of grouping studies according to common predictors and outcomes reduced heterogeneity in our meta-analysis, it resulted into a small number of studies being included in each analysis, which did not allow for an accurate analysis of risk of bias. As recommended by the Cochrane guidelines for meta-analyses (Higgins et al., 2019), approximately 10 studies are desirable in order to assess whether funnel plot asymmetries are driven by study characteristics rather than by chance, therefore funnel plots were not included for all studies. Nonetheless, our approach meant that the heterogeneity indexes yielded from our analysis were relatively low.

In light of the points discussed above, the use of homogeneous, more extensive testing protocols is encouraged in longitudinal research. These protocols should include approximately 5 tests targeting different aspects of cognition, which should be averaged to compute a global cognitive score, that would thus express global cognition in the older population with improved sensitivity. Adopting homogeneous testing guidelines for ageing research would also facilitate cross-study comparisons, by allowing for comprehensive meta-analysis, with low heterogeneity indexes. The use of wearable devices and computerised methods to assess physical function is also encouraged, in order to obtain objective, sensitive measures of physical function. Finally, the study of reciprocal associations between physical and cognitive functions is encouraged, in order to better understand the directionality of physical-cognitive functions associations. While some of these suggestions may seem more feasible than others, establishing solid guidelines in longitudinal research and adhering to consistent protocols may play a significant role in our understanding of the dynamic relationships between variables, which may be extremely important to promote optimal ageing.

Chapter 3: Longitudinal association between performance-based physical fitness measures and dementia risk: a systematic review of prospective studies

Introduction

This chapter will review longitudinal evidence on the association between performance-based physical fitness measures and dementia risk. As previously discussed, two already existing systematic reviews were identified looking at the prospective association between physical fitness and dementia risk. Hamer and colleagues' (2009) investigation included 16 studies looking at the association between physical activity levels and a later diagnosis of neurodegenerative disease. While Hamer et al.'s review shows an association between greater physical activity levels and reduced dementia risk, the studies included assessed physical activity through self-reported questionnaires, which introduces the issue of recall bias to the validity of the results. Blondell and colleagues' (2014) review looked at the association between physical activity and both cognitive decline and dementia risk outcomes, with 26 studies looking at the latter association. This review showed that greater levels of physical activity were associated with reduced risk of dementia at follow-up; however, once again this review relied almost exclusively on self-reported markers of physical activity. This chapter will review longitudinal evidence focusing on the association between objectively-assessed physical function and dementia diagnosis at follow-up, following from the search process that was outlined in the methods section of chapter 2.

Methods

After the process of study selection that was described in chapter 2, 50 studies were included in the review (refer to p. 28 for flow diagram illustrating study selection and description of the selection process). 35 of these studies had a cognitive decline outcome and were analysed and discussed in the previous review. The remaining 15 studies had a diagnostic outcome: that is to say, participants were categorised based on whether they had developed dementia at follow-up. This chapter will focus on data extraction, quality assessment and meta-analysis of the evidence considering a dementia diagnosis as the primary outcome measure. The decision of dividing the reviews granted consistency in the outcome measures, as well as in the statistical indexes of association between physical function and cognition, which allowed for more comprehensive and less heterogeneous meta-analyses. Data from the 15 studies with dementia as a primary outcome were extracted using a standardised form, including information on: first author and year of the study, country, study from which the data was obtained, participant demographics, length of study follow up, performance-based measures of physical function, cognitive or diagnostic outcomes, and primary findings from the study. The data extraction process is summarised in Table 3 below.

Data extraction

Table 3: Data extraction of relevant study variables, for studies with a dementia outcome

Author, year	Country	Study	Participants	Baseline age range	Gender (Female %)	Length of follow-up	PF measure	Cognitive measure	outcome
Boyle 2009	USA	the Rush Memory and Aging Project	970	> 54	75,15 %	3.6 years	muscle strength and handgrip strength	Alzheimer's disease diagnosis	grip strength was associated with the risk of AD
Buchman, 2007	USA	the Rush Memory and Aging Project	823	>65	69,4%	5.7 years	handgrip strength	Alzheimer's disease diagnosis	each 1-lb decrease in baseline grip strength was associated with a 1.5% increase in the risk of AD
Buchman, 2012	USA	the Rush Memory and Aging Project	716	> 54	76%	4 years	accelerometer data	Alzheimer's disease diagnosis	total daily physical activity was associated with incident AD
Doi, 2019	Japan	National Center for Geriatrics and Gerontology Study of Geriatric Syndromes	4086	≥ 65	52%	3.5 years	handgrip strength, sit-to-stand test, timed up and go test	dementia diagnosis	sit to stand test and timed up and go, but not handgrip strength, were associated with increased dementia risk
Dumurgier, 2017	France	The Three-City Study	3663	> 65	61.90%	9 years	walking speed	dementia diagnosis	lower walking speed was associated with an increased hazard of dementia
Hackett, 2018	UK	The English Longitudinal Study of Ageing (ELSA)	3932	> 60	56%	12 years	walking speed	dementia diagnosis	walking speed was a predictor of development of dementia
Lee, 2015	Hong Kong	data from the Elderly Health Centre of the Department of Health, Hong Kong	1775	>65	62, 47%	6 years	balance test	dementia diagnosis	poor balance was related to greater dementia risk
Lee, 2018	Korea	The Korean National Health Insurance Service-National Health Screening Cohort database	49283	> 66	50.50%	3.8 years	timed up and go test	dementia diagnosis	poor performance in TUG was related to a higher risk of total dementia incidence

Montero-Odasso, 2018	Canada	The Gait and Brain Study	154	> 65	54.50%	5 years	walking speed	dementia diagnosis	Slow gait at baseline failed to predict dementia
Sattler, 2011	Germany	The Interdisciplinary Longitudinal Study on Adult Development and Aging (ILSE)	381	> 59	49.40%	12 years	balance test, handgrip strength	dementia diagnosis	Better performance in balance tests at baseline reduced dementia risk, no associations were observed with handgrip strength
Sibbett, 2018	United Kingdom	The Lothian Birth Cohort 1921 (LBC1921)	488	> 65	57.40%	12 years	Forced expiratory volume (FEV1), handgrip strength, walking speed	dementia diagnosis	none of the physical function measures were found to be associated with dementia
Taniguchi, 2017	Japan	Data from a 13-year longitudinal study launched in Kusatsu Town in 2002	1686	>65	56,3%	7,22 years	walking speed	dementia diagnosis	participants in the low gait speed group had higher risk of developing dementia
Wang, 2006	USA	The Adult Changes in Thought (ACT) study	2288	≥ 65	60%	5.9 years	handgrip strength, balance test, sit-to-stand test, walking speed (PPF score)	dementia diagnosis	walking speed, balance, and grip strength were associated with reduced dementia risk
Welmer, 2014	Sweden	The Swedish National study on Aging and Care	2938	≥ 60	82.10%	6 years	walking speed	dementia diagnosis	each unit decrease in baseline walking speed increased the likelihood of incident dementia
Wilkins, 2013	USA	the Knight Alzheimer's Disease Research Center (ADRC)	435	≥ 60	60.90%	5 years	9-item Physical Performance Test (PPT)	dementia diagnosis	the physical function scores significantly predicted time to dementia diagnosis

Characteristics of the studies

15 studies were retrieved which examined the association between physical function at baseline and subsequent dementia diagnosis. The total number of participants from the studies included was 72902, with participants from all studies aged 50 years or older. The sample size of the studies included ranged from 154 to 49283. All of the studies included were prospective longitudinal cohort studies, with follow-up durations ranging from a minimum of 3.5 years to a maximum of 12 years. Most of the studies included were conducted in the USA (5 studies), followed by Japan (2 studies), United Kingdom (2 studies), and Korea, Hong Kong, Canada, Germany, Sweden and France with 1 study each. All the studies included both genders, with a mean female percentage of 60.1%. The studies included in the review employed diverse performance-based measures of physical function: 6 studies used handgrip strength, 7 studies employed walking speed, 1 study assessed accelerometer-derived daily step count, 3 studies used balance measures, 2 used the timed up and go test, one looked at cardiorespiratory measures of physical function, 2 used chair-stands test and one study employed the 9-item physical performance test. These measures were combined in some of the studies. The 15 studies included had dementia diagnosis as the primary outcome measure assessed at follow-up. All of the studies included covariates in the analyses, the most common of which were age, sex, education, co-morbid health conditions, BMI, depression, baseline cognition, and social engagement.

Results

Assessment of study quality and covariates

Quality assessment was conducted using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Cohort Studies (Moola et al., 2017), which was modified by adding additional questions on sample size, and on relevant covariates that should be included in longitudinal statistical analyses looking at the association under investigation. Based on whether studies met these criteria, a score ranging from 0 (corresponding to No or N/A), 1 (Unclear/Partially), or 2 (Yes) was attributed to each criterion. A total score was computed for each study: a score of 16-20 identified high-quality studies, a score of 11-15 identified studies of fair quality and a score of 5-10 identified studies of poor quality. The mean quality of the studies included was 16.40, with scores ranging from 14 to 20 and with an overall fair to high level of the studies assessed. 6 studies were defined as being of fair quality, and the remaining 9 were identified as high-quality studies. The quality assessment has been summarised in Table 4 below (see p. 71).

The studies included in this review used covariates that can be categorized into: (1) demographic factors, (2) social factors such as socioeconomic status and education, (3) lifestyle factors such as smoking or diet (4) comorbidities such as cardiovascular disease, hypertension or diabetes, (5) health status, such as depression, anxiety symptoms, body mass index, or cognitive status and (6) genetic factors such as *APOE4* status. While the significant associations between physical function and dementia risk was mostly unaffected by confounders, some of the studies reported that specific covariates attenuated this association. In particular, Dumurgier et al. (2017) and Welmer et al. (2014) reported that the association between physical function and dementia risk was attenuated by adjusting for baseline cognitive test performance, whilst Doi et al. (2019) found that comorbidities and baseline cognitive function were found to attenuate this association.

Table 4: Assessment of study quality

First author, year	Were the two groups similar and recruited from the same population?				Were the exposures measured similarly to assign people to both exposed and unexposed groups?				Was the exposure measured in a valid and reliable way?				Were confounding factors identified? (at least age, gender, education, comorbidities)				Were strategies to deal with confounding factors stated?				Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)?				Were the outcomes measured in a valid and reliable way?				Was the follow up time reported and sufficient to be long enough for outcomes to occur?				Was follow up complete, and if not, were the reasons to loss to follow up described and explored?				Were strategies to address incomplete follow up utilized?				Was appropriate statistical analysis used?				Was the sample size appropriate?				Score:
	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	Y	N	U/P	N/A	
Boyle 2009				0				0	2				2				2				2				2						1		2							0	2						1		16
Buchman, 2007				0				0	2				2				2				2				2				2				2							0	2						1		17
Buchman, 2012				0				0			1		2				2				2				2						**1**		2							0	2						1		15
Doi, 2019				0				0	2				2				2				2				2						1		2							0	2				2				17
Dumurgier, 2017				0				0	2				2				2				2				2				2				2				2				2				2				20
Hackett, 2018				0				0	2				2						1				1		2				2						1		2				2				2				17
Lee, 2015				0				0	2						1				1		2				2				2				2							0	2				2				16
Lee, 2018				0				0	2				2				2				2						1				1		2							0	2				2				16
Montero-Odasso, 2018				0				0	2				2				2				2				2					0			2							0	2					0			14
Sattler, 2011				0				0	2				2				2						1		2				2				2					0			2					0			15
Sibbett, 2018				0				0	2				2				2				2						1		2				2							0	2					0			15
Taniguchi, 2017				0				0	2				2				2				2						1				1		2				2				2				2				18
Wang, 2006				0				0	2				2				2				2				2				2				2				2				2				2				20
Welmer, 2014				0				0	2				2				2				2				2						1		2					0			2					0			15
Wilkins, 2013				0				0			1		2				2				2				2				2				2					0			2					0			15

Y = Yes, N = No, U/P = Unclear/ Partially, N/A = Not applicable

Statistical analysis

Different meta-analyses were conducted to study the correlation between performance-based physical function and dementia risk, where pooled mean estimates were computed using the review and meta-analysis software RevMan review manager 5.4.1. Random-effect models were employed to calculate pooled effect sizes from Odds/Rate/Hazard Ratios and 95% confidence intervals, in order to account for the heterogeneity of follow up periods and sample sizes. Studies were grouped into different meta-analyses according to common physical function predictors. The physical function tests that defined the predictors under investigation in this meta-analysis were defined as walking speed, handgrip strength, sit-to-stand test, balance and timed up and go test, while the outcome measure of interest was dementia diagnosis, which characterised all of the studies included. This resulted in 5 meta-analyses, one for each of the physical function predictors. Publication bias was examined using funnel plots, which were produced using RevMan 5.4.1, where at least 5 studies were deemed necessary to assess risk of bias. Heterogeneity between the studies included in each meta-analysis was assessed via the I^2 statistic and the associated p value, where values between 25-50% identified low heterogeneity, 50-75% indicated moderate heterogeneity, and values greater than 75% identified high heterogeneity. Where heterogeneity was high, the analysis was followed by sensitivity analysis which excluded individual studies from each meta-analysis, to investigate the influence of single studies on mean estimates.

Meta-analysis

Handgrip strength and dementia risk

A total of 6 studies looked at the association between handgrip strength and dementia risk. Of these studies, only Sattler et al. (2011) expressed their results as a measure of muscle strength derived from complex strength measures covering the whole body (among which handgrip strength), while the rest of the studies relied solely on handgrip strength. Therefore, the findings from Sattler et al. (2011), which did not find composite muscular strength to significantly predict later dementia diagnosis, were not included in this meta-analysis. The results from this meta-analysis showed that handgrip strength was significantly associated with reduced dementia risk at follow up, HR= .94, 95% CI: .88–1.00, p = .050. The analysis was characterised by high heterogeneity, $I^2 = 76\%$, p = .002. Sensitivity analyses looking at the results of the meta-analysis when removing each study individually showed that when removing Boyle et al. (2009) from the meta-analysis, the heterogeneity was greatly reduced, $I^2 = 26\%$, p = .25, however the model was no longer significant (p = .120). The substantial contribution to heterogeneity may be driven by participants in Boyle and colleagues' (2009) investigation being younger (aged 54 or older) than the other studies included in the analysis (participants aged 65 or older for all the remaining studies), which may be partially inflating the magnitude of the observed association. The funnel plot for the analysis (See Figure 13) shows a fairly symmetric distribution of the studies, suggesting the absence of publication bias.

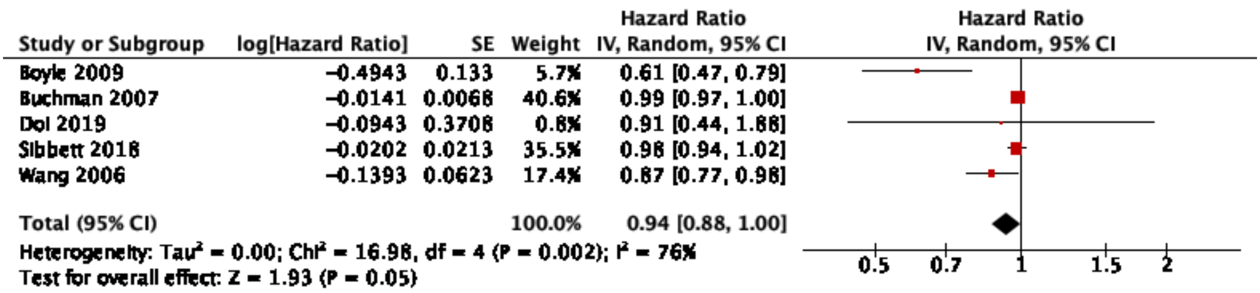

Figure 12: Forest plot for the association between handgrip strength and dementia risk

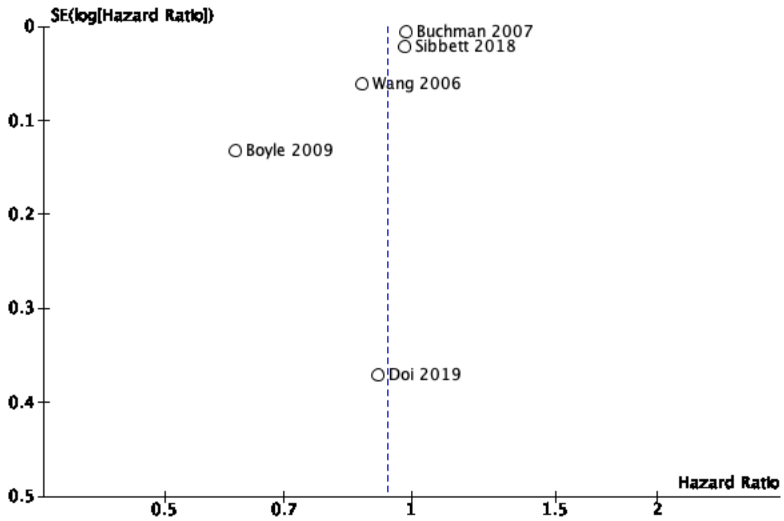

Figure 13: Funnel plot for the association between handgrip strength and dementia risk

Walking speed and dementia risk

7 studies were retrieved looking at the association between walking speed and dementia risk. Of these studies, only Welmer et al. (2014) expressed their results as ORs, while the remainder of studies used HRs. Therefore, findings from Welmer et al. (2014) showing that each standard deviation slower walking speed at baseline increased the likelihood of developing dementia were not included in the meta-analysis. The remaining 6 studies were divided into 2 further meta-analyses, given that 3 of them expressed their results in terms of units decrease in walking speed, while the other three expressed them as a comparison between slow or fast walking speed groups at baseline.

The results from the first meta-analysis showed that faster walking speed was significantly associated with reduced dementia risk at follow up, HR= .73, 95% CI: .54–1.01, p = .050. The analysis was characterised by high heterogeneity, $I^2 = 89\%$, p <.001, which was not reduced in further sensitivity analyses removing each of the studies from the meta-analysis individually.

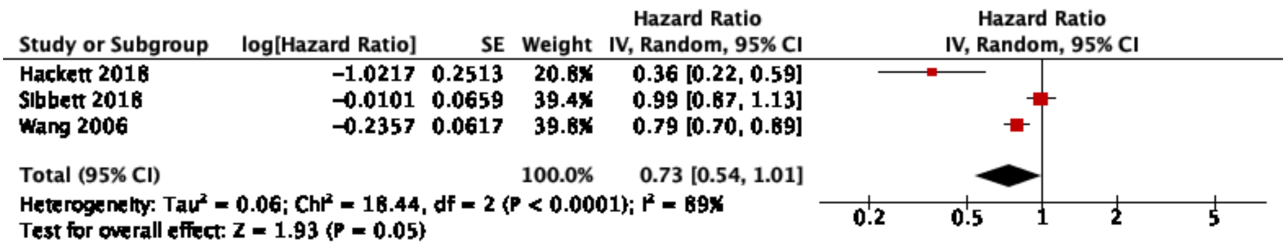

Figure 14: Forest plot for the association between walking speed and dementia risk

The results from the second meta-analysis showed that slower walking speed at baseline was associated with greater dementia risk at follow-up, HR= 1.59, 95% CI:1.39–1.81, p < .001 The analysis did not show heterogeneity, $I^2 = 0\%$, p = .83.

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
Dumurgier 2017	0.4637	0.0686	93.4%	1.59 [1.39, 1.82]
Montero-Odasso 2018	0.1484	0.5561	1.4%	1.16 [0.39, 3.45]
Taniguchi 2017	0.5306	0.2916	5.2%	1.70 [0.96, 3.01]
Total (95% CI)			100.0%	1.59 [1.39, 1.81]

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.37$, df = 2 (P = 0.83); $I^2 = 0\%$
Test for overall effect: Z = 6.98 (P < 0.00001)

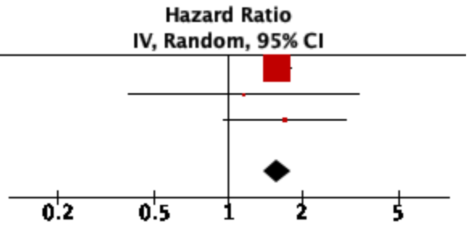

Figure 15: Forest plot for the association between walking speed and dementia risk

Balance and dementia risk

3 studies looking at the association between balance and dementia risk were identified from our search. Lee et al. (2015) was not included in the meta-analysis, because they expressed their results in terms of dementia likelihood in relation to poor balance performance, while Sattler et al. (2011) and Wang et al., (2006) expressed their results in relation to greater balance performance at baseline. Nonetheless, results from Lee et al. (2015) showed greater likelihood of developing dementia for participants with poor balance at baseline.

The results from the meta-analysis showed that high balance performance at baseline was significantly associated with lower risk of developing dementia at follow-up, OR= .85, 95% CI:.76–.94, p = .002. The analysis had high heterogeneity, $I^2 = 88\%$, p = .004; however, due to the number of studies in this meta-analysis, sensitivity analyses could not be carried out.

Study or Subgroup	log[Odds Ratio]	SE	Weight	Odds Ratio IV, Fixed, 95% CI
Sattler 2011	−1.0498	0.3117	3.1%	0.35 [0.19, 0.64]
Wang 2006	−0.1393	0.0557	96.9%	0.87 [0.78, 0.97]
Total (95% CI)			100.0%	0.85 [0.76, 0.94]

Heterogeneity: $Chi^2 = 8.27$, df = 1 (P = 0.004); $I^2 = 88\%$
Test for overall effect: Z = 3.05 (P = 0.002)

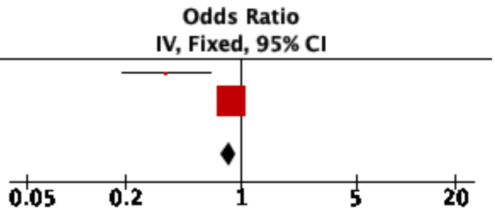

Figure 16: Forest plot for the association between balance and dementia risk

Timed up and go test and dementia risk

Two studies were retrieved looking at the association between performance in the timed up and go test and dementia risk. Results from this meta-analysis showed that poor performance in the timed up and go test at baseline was associated with greater risk of developing dementia at follow-up, HR= 1.36, 95% CI: 1.17–1.59, $p < .001$. The analysis did not show heterogeneity, $I^2 = 0\%$, $p = .55$.

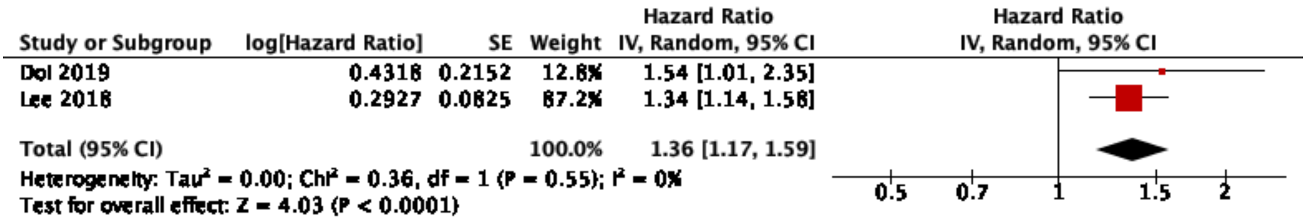

Figure 17: Forest plot for the association between timed up and go test and dementia risk

Chair-stands test and dementia risk

Two studies looking at the association between the chair-stands test and dementia risk were identified from our search. Results from this meta-analysis showed that higher performance in chair-stands test at baseline is associated with reduced risk of developing dementia at follow-up, HR= .86, 95% CI: .79–.94, $p < .001$. The analysis did not show heterogeneity, $I^2 = 0\%$, $p = .63$.

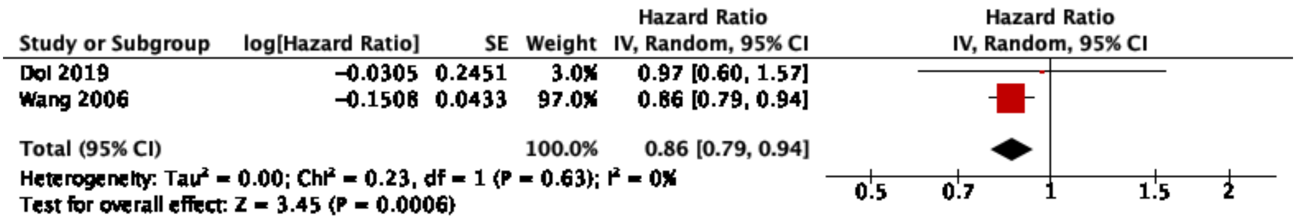

Figure 18: Forest plot for the association between chair-stands test and dementia risk

Synthesis of additional studies

Two studies (Buchman et al., 2012; Wilkins et al., 2013) were pertinent to the scope of this review, but were not included in the meta-analyses due to the use of specific predictors that were not comparable with the predictors that were analysed thus far. These predictors were accelerometer-based measures of daily life activity for one study (Buchman et al., 2012), and a composite physical performance test (PPT) score for the other study (Wilkins et al., 2013). These predictors were not analogous to the predictors that were grouped into the meta-analyses reported above; nonetheless, these studies offered valuable insight on the comparison of interest, therefore their results will be discussed narratively.

Buchman and colleagues (2012) used data from a 10-day period of accelerometer wearing to derive daily physical activity measures, and investigated the relation of the latter to the risk of developing dementia over a 4-year follow up: their results showed that greater daily physical activity levels were related with reduced risk of developing dementia at follow-up, HR = .47; 95% CI = .27–.83. On the other hand, Wilkins et al. (2013) used scores in the PPT, which is derived from daily life activity physical tasks such as simulated eating or dressing, combined with chair rises and a standing balance test to predict dementia risk. Wilkins and colleagues (2013) showed that worse performance in the PPT significantly predicted time to a dementia diagnosis, HR = .89; 95% CI = .86–.93. While these studies employed different methods to investigate the outcome of interest, their results both seem to point towards an association between physical function and dementia risk. Interestingly, the physical function measures that were described in both studies both capture aspects of daily life physical functioning. Indeed, accelerometer-based measures provide an objective indication of physical activity which captures low to moderate activities of daily life, while the PPT focuses on specific tasks that are involved in daily life household activities. Therefore, findings from these studies may be informative on the significance of daily life physical activities as a predictor of dementia risk.

Further to the measures that were employed in these studies, Sibbett and colleagues' (2018) investigation was the only study retrieved from our search that focused on lung function as assessed through forced expiratory volume (FEV1) as a predictor of dementia risk. Results from their study showed that, in contrast with previous evidence (Russ et al., 2015), lung function did not significantly predict dementia. The authors propose that their findings contradict existing literature due to previous studies investigating younger populations, with wider age ranges than the Lothian Birth Cohort 1921, where the mean age was 80.3 years old (Sibbett et al., 2018). Further, the authors did find that better lung function at age 79 was correlated with reduced risk of death, which may suggest that participants with poorer lung function, who would have been at higher risk of developing dementia, were more likely to have died of other causes prior to the onset of the disorder. When considering the high correlation between poor lung functioning and disorders such as diabetes or coronary heart disease (Pathan et al., 2011), the death of participants with poor lung functioning prior to the onset of dementia appears to be a reasonable explanation for these results. Both of the factors discussed by the authors may explain why the association between respiratory physical function and dementia risk, which has been established by previous research (Russ et al., 2015), was not supported by data from this study.

As far as the studies that were not included in the meta-analyses due to non-comparable statistical indexes of association, findings from Welmer and colleagues (2014) and Lee et al. (2015) both supported the results from our meta-analysis, showing increased risk of dementia as predicted by poor walking speed and worse balance performance respectively. On the other hand, findings from Sattler et al. (2011), did not find poor muscular strength as measured through a composite strength measure derived from muscle strength measures from the whole body (including handgrip strength, among other muscle groups) to significantly predict greater dementia risk. While previous studies have found muscular strength as assessed through handgrip strength to significantly predict MCI and Alzheimer's disease (Boyle et al., 2009), which were also supported by the results from our meta-analysis, the authors propose that the use of a composite measure as opposed to a measure of strength relying solely on handgrip strength may explain the differential results. This may also suggest that handgrip strength may be a more sensitive predictor of cognitive impairment compared to measures relying on several muscle groups. Recent studies have proposed that the greater involvement of cognitive processes that are required for hand dexterity, such as planning and attention, which are also subject to age-related deterioration, may explain its greater role in predicting cognitive performance and dementia risk (Kobayashi-Cuya et al., 2018). Therefore, results from Sattler et al. (2011), when examined in light of these considerations and in view of the findings from our meta-analysis on handgrip strength, seem to point towards the preferable use of handgrip strength as an optimal predictor of dementia risk.

Discussion

This chapter investigated the predictive role of objective physical function for dementia risk, by reviewing prospective cohort studies investigating the association between performance-based measures of physical fitness at baseline and dementia diagnosis at follow-up. The results from our meta-analyses indicated that poor performance in handgrip strength, walking speed, balance, timed up and go and chair-stands tests were all significant predictors of subsequent dementia diagnosis. Further findings from studies that were not included in the meta-analyses due to non-comparable predictors suggested that accelerometer-based daily physical activity and a composite measure derived from daily life physical activities were also predictive of dementia risk. These results are in line with previous research showing reciprocal associations between performance-based physical function measures and dementia incidence (Hamer et al., 2009; Blondell et al., 2014). In particular, a recent systematic review of longitudinal cohort studies focusing exclusvely on studies using handgrip strength as a predictor has shown that poor handgrip strength is associated with increased risk of cognitive decline and dementia (Cui et al., 2021). Similarly, meta-analytical data focusing on the predictive role of gait speed has found associations between low walking speed and dementia incidence (Beauchet et al., 2016). The associations between poor balance, chair-stands and timed up and go tests performance and dementia risk that we observed also appear to corroborate further literature documenting the predictive role of these physical function measures on dementia incidence (Bullain et al., 2016; Katsumata et al., 2011). Our review considered all of the measures of physical function discussed above, by gathering extensive evidence on a variety of objective predictors, instead of focusing solely on one measure. Our results for each of the predictors that we considered are in line with findings from reviews focusing exclusively on the same predictors. At the same time, our review includes studies which looked at a combination of different predictors, and pooled them into different meta-analyses according to common predictors, allowing for a comprehensive review of the literature.

Potential mechanisms

The association between physical functioning and dementia incidence is mostly mediated by the same mechanisms which promote the cognitive benefits associated with greater physical fitness levels. In addition, some of the primary hallmarks of Alzheimer’s disease seem to be significantly impacted by physical exercise. As discussed in the above review, skeletal muscle function facilitates the release of BDNF, which is known to mediate neuronal growth and plasticity, as well as learning processes. Interestingly, in-vitro studies have shown that BDNF is involved in the survival of neurons affected by neurodegenerative disorders such as dementia and Alzheimer's disease (Murer et al., 2001). Further, physical exercise has been shown to enhance dendritic length and density within the hippocampus, as well as promoting neurogenesis in this region (Eadie et al., 2005). Given the substantial hippocampal degeneration underlying the loss of memory that characterizes dementia and Alzheimer's disease, the role of physical exercise in promoting hippocampal neurogenesis may be defining its role as a protective factor for dementia. Additionally, physical exercise has been shown to affect the development of brain β-amyloid and TAU protein. β-amyloid and TAU protein are primary markers of Alzheimer's disease, the most frequent type of dementia (Ahlskog et al., 2011). While the former is the main component of neuritic plaques, the latter makes up neurofibrillary tangles: both of these factors are believed to be the primary cause underlying the cognitive deterioration typical of Alzheimer's disease. Intervention studies on animals have shown that a 9-month exercise program may prevent the development of TAU protein in the hippocampus (Belarbi et al., 2011). Further evidence has shown that greater levels of long-term exercise are correlated with reduced brain accumulation of β-amyloid (Liang et al., 2010). Therefore, optimal exercise levels may contribute to reduce the risk of developing Alzheimer's disease. Further, physical exercise is known to promote optimal blood flow in the brain, which may reduce dementia risk by mitigating vascular risk factors for dementia such as cardiovascular brain damage or small vessel disease (Ahlskog et al., 2011).

Altogether, the findings from our meta-analysis indicate a protective role of optimal physical function against dementia risk. Given the growing rates of the ageing population, and the known threat that dementia represents for our healthcare and societal systems, these results have significant implications for clinical practice and policy making. For instance, knowing that poor physical functioning predicts dementia risk may encourage the clinical monitoring of physical functioning starting in midlife. Adopting this approach in standard clinical practice could contribute towards the identification of potential early markers of dementia, which may later degenerate into a dementia diagnosis. At the same time, the knowledge that optimal physical function decreases the risk of dementia may inform the development of intervention programmes, where specific physical functions can be trained for a set period of time. Interventions may not only promote greater overall physical functioning, but also facilitate the mediating mechanisms discussed above, with the potential end result of improving cognitive trajectories.

Our study had some limitations. While our analysis investigated the chronological association between physical function and dementia risk, it did not examine their causal relation. That is to say, our results do not inform us on whether poor physical performance leads to greater dementia risk, or if, conversely, poor cognitive functioning may be causing worse physical performance. Indeed, prospective cohort studies are informative on associations between two variables, but by way of their observational nature, it is not possible to exclude the existence of reverse causality between the variables under investigation. Therefore, it is possible that low physical fitness levels may be an actual risk factor for dementia, through the facilitation of BDNF release and the attenuation of cardiovascular brain damage described above. However, at the same time, it is possible that people with dementia may experience physical slowing, along with cognitive slowing, as part of the early stages of dementia which precede its diagnosis (Andrade, 2020). In this case, dementia would be the cause for poor physical functioning. Our study partially mitigated for the issue of reverse causality, by only including studies looking at participants with no cognitive impairment at baseline. Although cognitive changes may be latent in prodromal dementia, this strategy excluded the possibility of participants being affected by MCI or evident cognitive impairment at baseline. In turn, the exclusion of participants with cognitive decline at baseline indicates that poor physical functioning was not attributable to physical slowing in early stages of dementia. Another option is that the presence of cardiovascular disease may be a common cause predisposing those affected to both poor physical functioning and dementia. However, most of the studies that we included featured high-quality statistical analyses, controlling for the mediating effects of comorbidities.

One issue emerging from this investigation is that of inconsistencies in the reporting of results. This is exemplified by the meta-analysis looking at the association between walking speed and dementia risk, which had to be further separated into two meta-analyses, with a lower number of studies each. A further division was necessary because out of the studies looking at the association of interest, 3 of them expressed their results using the highest-performing group as a reference, while the other 3 used the lowest-performing group as reference. This resulted in the former reporting lower risk of dementia at follow-up as predicted by greater walking speed performance, while the latter studies reported greater dementia incidence as preceded by poor walking speed. Although these results are consistent with the performance reference adopted by each study for walking speed, and both sets of results suggest a protective effect of greater walking speed against dementia risk, the inconsistency in reporting generates non-comparable statistical indexes of association. The adherence to consistent reporting guidelines is encouraged, and would allow for comprehensive meta-analyses of the variables of interest. For instance, our search also retrieved studies such as Doi et al. (2019) which extensively reported their results, including frames of reference for both low and high-performing groups for each physical variable under investigation: this reporting method makes cross-study comparisons possible, and facilitates a comprehensive meta-analytical approach.

Our study attempted to minimise heterogeneity in the meta-analyses, by grouping studies according to common predictors. This strategy was adopted in light of the high heterogeneity that characterised our pool of studies, in view of great variability in sample sizes, age range of participants, and length of follow-up periods. While this approach allowed for more coherent analyses where predictors were homogeneous, it also resulted in the number of studies included in each meta-analysis being relatively small. This made the assessment of risk of bias and the conduction of sensitivity analyses problematic, or not feasible in some cases. Our quality assessment tool partially mitigated for the lack of funnel plots for some of the meta-analyses, by assessing the adherence of the studies to pre-established quality standards. These quality standards also included the adjustment of statistical analyses for age, gender, education and comorbidities, which have been identified as potential confounders in the association between physical function and dementia risk. While most of the studies included in this review controlled for an extensive range of covariates in their analyses, the quality assessment process identified one study (Lee et al., 2015) which failed to adjust their analysis for confounders that are known to potentially influence the association between physical function and dementia, such as education and comorbidities. The inappropriate adjustment for confounders in statistical analyses may strengthen the magnitude of the association between physical function and dementia, and potentially influence the final findings. The adherence to consistent covariate adjustment in statistical analyses is thus strongly encouraged in longitudinal research.

Another problematic aspect of grouping studies into different meta-analyses according to common predictors is that where only one study was retrieved looking at the association between a specific predictor and dementia, the results from said study could only be discussed narratively. In this review, Buchman et al. (2012) and Wilkins and colleagues (2013) found that accelerometer-based measures of daily physical activity for the former, and a composite physical performance test (PPT) score for the latter were both significant predictors of dementia risk. As discussed above, the physical function measures that were used in these studies measure aspects of daily life physical functioning. While accelerometer measures of daily physical activity provide an objective indication of daily steps, which includes low to moderate activities carried throughout the day, the PPT measures specific tasks that are involved in daily life activities, revolving around writing, eating, getting dressed, etc. Findings from these studies may be informing us on the relevance of the low-to-moderate physical functioning involved daily life activities to predict cognitive outcomes. The use of composite measures, as well as the use of devices that may capture low levels of physical activity which may not be effectively encompassed by the other performance-based measures that we investigated in this study, is therefore strongly encouraged.

In conclusion, our study lends support to the association between physical functioning and dementia risk. Results from our meta-analyses suggested that handgrip strength, walking speed, balance, chair-stands and timed up and go tests were all significant predictors of dementia risk. We also reviewed evidence suggesting that accelerometer-based and composite measures of physical function may predict dementia incidence, although this evidence was not included in the meta-analyses. We have identified some methodological limitations, such as the issue of reverse causality, as well as inconsistencies in the reporting of results, and heterogeneous adjustment for covariates, which may be inflating the associations that we observed. However, the quality assessment tool that we used, as well as our inclusion criteria ensured that these problematic aspects were partially mitigated for. The association that we observed is significant for clinical practice and policy making, and has considerable implications for both of these sectors, in view of the rapidly ageing population.

Chapter 4: Performance-based and self-reported measures of physical fitness in ageing research

Introduction

This chapter will examine the primary challenges of measuring physical fitness levels in ageing research. As previously discussed, being able to measure physical fitness in older adults is central to investigating the predictive role of this variable on cognitive status, with significant implications for the design and implementation of public health policies. The predictive value of physical fitness for several aspects of health and cognition makes it a salient modifiable risk factor for adverse health and cognitive outcomes, with the remarkable potential to influence these trajectories. Additionally, the precise measurement of physical performance allows us to accurately monitor changes in physical fitness levels, which is highly significant when attempting to measure the effectiveness of intervention programmes targeting this variable (Prince et al., 2008).

As previously discussed in chapter 1, physical activity levels are frequently assessed through self-report questionnaires in epidemiological research. Self-report questionnaires are a practical, easily administrable way to measure physical fitness in the elderly population (Ainsworth, 2009; Innerd et al., 2015). In these questionnaires, participants are asked to report their levels of engagement in physical activity, either through a single score which represents overall physical activity, or by reporting the number of minutes spent on different activities over a specific timespan (Hamer et al., 2009; Sofi et al., 2010; Blondell et al., 2014; Watkinson et al., 2010). While self-report measures capturing physical activity in the elderly are broadly accurate and reliable, there are several factors which undermine their validity, particularly when considering the high prevalence of cognitive deficits in the target population. This chapter will investigate some of these problematic aspects, as well as the current understanding on the consistency between self-report and objective measures of physical fitness, and discuss some of the challenges of employing objective measures of physical fitness in ageing research. A secondary data analysis will then be conducted, using data gathered from three cross-sectional studies conducted between 2017 and 2019, which looked at the correlation between physical fitness and cognition. These studies used both self-report and performance-based measures to assess physical fitness. The agreement between these measures will be analysed using correlation analyses, in order to investigate the consistency between objective and subjective outcomes of physical fitness. Further, a regression analysis will look at the contribution of age, and subjective and objective measures of physical fitness in predicting cognitive function as assessed by the MoCA. Based on the concept on convergent validity which will be further elaborated in this chapter, if subjective and objective measures of physical fitness have good agreement and are measuring the same construct, which has well-documented links with cognition, then both measures should explain a significant amount of variance in the cognitive outcome.

Evaluating a complex construct

Physical fitness may be defined as a set of characteristics that relates to the ability to perform physical activity (Hollamby et al., 2017). Physical activity is defined as any movement produced by our body, which results in the consumption of energy (Caspersen et al., 1985). Different types of bodily movement, which define different domains of physical activity, may make up for this energy expenditure (Trost & O'Neil, 2013). Examples of physical activity domains are leisure movement, sports, household work, walking, occupational physical activity or structured physical exercise. Furthermore, a single type of physical exercise may be characterised by different dimensions such as intensity, duration and frequency (Trost & O'Neil, 2013). As a result, the complex and diverse set of behaviours that compose physical activity, which in turn relates to the concept of physical fitness, lead to a range of biological responses from the body, which may be reflected by different performance-related features. These features may vary according to the domain (leisure movement, structured exercise, etc) and the dimension (intensity, duration and frequency) of the activity in question. Additionally, physical activity involves the synchronous collaboration of diverse physical functions, namely muscular strength, muscular endurance, and cardiorespiratory endurance (Caspersen et al., 1985), which further contribute to the complexity of evaluating this construct.

Due to the multifaceted nature of physical fitness, it is difficult to fully encompass the properties that make up this variable. Two predominant measurement strategies have been identified, which attempt to capture the complexity of this variable: these are either self-report questionnaires or objective measures of physical functioning. Self-report is very popular in ageing research due to their great practical appeal, cost-effectiveness and feasibility (Dishman et al., 2001). However, this strategy may not be the best way to assess physical fitness levels in the older population, as questionnaires rely solely on self-report, which may be biased by recall deficits, as will be further discussed below. Alternatively, objective measures of physical activity provide a real-time, performance-based indication of this variable. They may rely on observation of performance, such as with chair stands, walking speed or balance tests, or utilise technology to measure physiological parameters, such as in the case of accelerometer-based measures, heart rate monitoring, or grip strength, which are measurable through the use of different wearable devices. The direct observation of performance, as well as the reliance on technology, eliminate the recall bias that is typically associated with self-report measures.

Issues with the use of self-report measures in ageing research

While the use of objective measures of physical fitness mitigates the issues associated with self-reporting, self-rated measures are still predominantly employed in ageing research. Epidemiological studies make use of a range of methods such as interviews, questionnaires and activity log sheets to assess physical fitness levels. While these measures have great practical appeal due to their cost-effectiveness, easy administration and moderate burden on participants (Dishman et al., 2001), there are several issues associated with relying solely on such measures. In physical activity questionnaires, participants are prompted to report their physical activity levels, either by providing a single score or by reporting the minutes spent doing different physical activities over a specific time span, such as a week or a month (Sofi et al., 2010; Blondell et al., 2014; Watkinson et al., 2010). These questionnaires need to be designed with caution, in order to mitigate recall bias, which is very common in ageing research due to the high rates of cognitive decline in the older population (Shepard, 2003; Innerd et al., 2015). Commonly, the accuracy of self-report information tends to be differently affected by recall bias depending on the length of the time span that participants are asked about, and the complexity of the activities that they address (Dishman et al., 2001). Questionnaires enquiring about longer time spans are likely to capture participants' habits more accurately, but at the same time they involve a greater extent of recollection from memory. Another problematic aspect of response bias is represented by social desirability, whereby questionnaire respondents may answer questions in a way that is perceived favourably by the researcher: this may lead to participants overestimating their activity levels. Physical activity questionnaires designed for the older population should also effectively capture a range of physical activities that are meaningful for the elderly, which should reflect the high rates of limitations in physical capability, as well as the functional dependency that characterise this population (Shepard, 2003**).**

Questionnaires may also compromise the validity of the outcomes measured due to the nature of self-report, which is subject to biases of under and over-estimation of the actual engagement in physical activity (Parker et al., 2008). Indeed, there is evidence showing that when self-reporting physical activity levels, individuals might under-estimate how many minutes they spend participating in moderate activities (Ainsworth et al., 2000: Schaller et al., 2015). At the other extreme of this phenomenon, it is estimated that nearly 60% of adults who do not respect recommended guidelines for physical activity tend to overestimate their levels of activity (Van Sluijs et al., 2007). Additionally, studies show that individuals who have favourable indicators of physical fitness, such as lower Body Mass Index (BMI), tend to overestimate their activity levels (Watkinson et al., 2010). Cases of over and underestimation may provide an inaccurate picture of physical functioning (Parker et al., 2008):

such underestimations might mitigate the strength of the association between physical activity measures and cognition, whereas self-reported over-estimation might inflate this correlation. Finally, there are issues with the validity and reliability of subjective measures of physical fitness. In fact, a recent review on the measurement qualities of self-report physical fitness questionnaires has shown that self-report measures lack proof of validity (Helmerhorst et al., 2012). This review included 96 studies investigating the validity of both newly developed and already existing physical activity questionnaires, using accelerometer-based measures of physical activity to investigate validity. The measurement properties of 130 questionnaires were examined, including widely used measures such as IPAQ and PASE which were described in chapter 2. The validity of the investigated questionnaires ranged from .30-.39 for existing questionnaires, and from .25-.41 for newly developed questionnaires, which is considered to be poor to moderate validity. The pooled Spearman r for IPAQ validity was .33, 95% CI .26-.33, with values ranging from .27-.61, indicating modest rates of validity. This pattern of results suggests that the ability of self-reported measures of physical activity to measure this construct is still poor and, as highlighted by the authors, indicates the difficulty of capturing the greatly variable set of behaviours that characterise physical activity by relying solely on self-report (Helmerhorst et al., 2012).

On the other hand, objective measures of physical fitness do not rely on self-report. Both performance-based and device-measured markers of physical function provide an objective indication of performance, and they offer many advantages over self-report measures, as they are free from recall bias and over and underreporting issues. Additionally, the employment of performance-based measures ensures a comprehensive assessment, which can capture numerous aspects of a multifaceted concept. As previously discussed, physical functioning involves various factors which are related to both the skill and the health status of the individual: it is composed by muscular strength, muscular endurance, and cardiorespiratory endurance (Caspersen et al., 1985). These are respectively the amount of force a muscle can produce with a single effort, the ability to sustain muscle contraction for a relatively long period of time, and the ability to supply oxygen and other essential nutrients to working muscles. Research focusing on the associations between physical functioning and cognition should include objective measures that encompass at least one of these components of physical functioning, in order to provide an accurate depiction of physical status.

Objective measures of physical fitness have been available and employed in research for nearly 40 years; they primarily include grip strength as measured through dynamometer, walking speed, standing balance, chair rises, heart rate monitoring, and cardiorespiratory markers. These measures offer an accurate indication of the physical status of the individual, which objectively captures at least one of the dimensions that make up the complexity of physical functioning. Further, objective measures are not subject to recall bias, or under and overestimation biases, as well as being highly reliable and reproducible. The use of wearable devices to objectively assess physical activity is also becoming increasingly popular in epidemiological research, partially as a practical response to the self-report bias introduced by questionnaires. Instruments such as accelerometers and pedometers are easily accessible, cheap and reliable devices to objectively measure physical activity (Sesso, 2007; Parker et al., 2008).

The paucity of objective measures of physical fitness has been stressed by current epidemiological research. In a recent systematic review of prospective studies investigating longitudinal associations between physical fitness levels and cognitive decline in later life, Blondell and colleagues (2014) retrieved only two sources assessing physical fitness through an objective measure. Consequently, the authors highlighted the need for the increased use of objective measures of physical fitness in longitudinal cohort studies, to ensure a reliable investigation of physical function-cognition associations (Blondell et al., 2014). It is of note that Blondell and colleagues' search strategy may not have used search terms that were appropriate to retrieve objective evidence on physical function. Indeed, the authors reported using the search terms "physical activity" OR "exercise". Conversely, the majority of evidence focusing on objective predictors of physical function is retrieved using "physical function" or "physical functioning" as search terms, as we used for our search, and as was the search strategy in Clouston et al.'s (2013) review. Nonetheless, our review of the literature, as well as existing reviews (Hamer et al., 2009; Sofi et al., 2010) are sufficient to document the widely predominant use of self-reported measures of physical fitness in epidemiological research.

The use of different search terms, generating such different results raises the question of whether the terminology used to refer to physical fitness may be poorly specified. As previously discussed, physical activity is referred to as any bodily movement resulting in energy expenditure. Physical fitness refers to the capability to perform life activities with optimal performance, endurance and strength, and it constitutes characteristics that relate to the capability to perform physical activity. Physical function refers to functions such as mobility or dexterity, which define one's ability to perform daily life activities. In the literature that we surveyed, the term "physical activity" is usually used in relation to self-reported questionnaires, as these questionnaires usually enquire on and define, in fact, levels of engagement in different *physical activities*. Therefore, a search strategy relying exclusively on the term "physical activity" is bound to generate results including mostly self-reported measures of physical activity. Conversely, the search terms "physical fitness" and "physical function" will predominantly generate results relating to performance-based measures such as mobility, endurance, and strength, by way of the definition of these terms which refers to capability and performance-based dimensions. In light of the search terms used being able to generate significantly different results, the difference between these terms should be acknowledged by future research, in order to conduct searches that generate results in line with the scope of a specific investigation.

While the transition to a performance-based approach is highly desirable, it is still far from reality. As a matter of fact, in spite of the problematic aspects that are related to the intrinsic nature of self-report, questionnaires are still predominantly used, as in many circumstances they are the most practical, cost-effective way to assess physical function. Further, objective measures of physical performance do not come without their own limitations. High quality monitoring devices are costly, place greater burden on participants, and require trained staff to be accurately used. For this reason, these measures are mostly used in small-scale studies as opposed to larger prospective cohort trials, which was reflected by the sample sizes of the studies that we retrieved in chapters 2 and 3, where accelerometers were used. Ultimately, the most suitable method to assess physical fitness depends on a variety of factors such as sample size, length of the study, and funds available for the study, which should all be considered when developing study protocols.

Secondary data analysis of objective and self-report physical fitness data of healthy individuals

Consistency between objective and self-reported measures of physical fitness: existing evidence

As discussed thus far, there is still no consensus in regards to the gold standard for the optimal measurement of physical fitness in older adults. While the debate is still ongoing, the output from questionnaires on physical activity is often compared with performance-based measures, in order to establish whether the output generated from these measures reflects similar results. As a matter of fact, there is a growing body of cross-sectional research assessing physical fitness through a combination of self-report methods and performance-based measures, in order to investigate the agreement between these variables and determine whether their outcomes are consistent. The presence of cross-sectional associations between self-reported and performance-based measures of physical function substantiates the validity of self-reported measures and lends credibility to their use in ageing research, as it demonstrates that self-reported measures may be appropriately capturing the construct of interest, which should be best encompassed by performance-based markers.

In a recent systematic review, Prince and colleagues (2008) reviewed evidence from 187 studies looking at the consistency between self-report physical activity as measured through questionnaires and diaries, and objective measures of physical performance. The predominantly used direct measures of physical activity were accelerometer and double-labelled water, a method used to estimate energy expenditure by examining the difference between the turnover rates of hydrogen and oxygen of body water. Various self-report measures were used to assess physical activity levels. The most common self-report measure was the Seven-Day Physical Activity Recall (7-day PAR), which provides a self-report estimate of the time spent on physical activity, flexibility and strength tasks in the 7 days prior to testing (Sallis et al., 1985). Other frequently used tests were variations in duration of the 7-day PAR, such as the 24-hr PAR or the 4-week PAR, or the IPAQ. The authors' meta-analysis showed low-to-moderate agreement between self-report and objective measures of physical function, and found no clear trend in the association between these variables, with self-reported measures of physical activity being both higher and lower than values measured with direct methods. For instance, self-report measures estimated higher levels of activity than accelerometers, while subjective measures estimated lower levels of activity than those directly shown by double labelled water. As argued in this extensive review, the lack of strong correlations or evident patterns of agreement between objective and self-report variables is a significant cause for concern when these tests are employed interchangeably (Prince et al., 2008). The review highlighted that the use of accelerometer measures as an objective test by many of the studies included in this review may have been problematic, as these tests are unable to effectively capture the energy expenditure from activities using upper limbs (Prince et al., 2008). Further, the results highlighted that self-report measures were generally unable to accurately capture greater levels of physical activity, as they appeared to over-estimate levels that are reported as “vigorous” by questionnaire respondents. Moreover, the authors highlighted that by way of their structure, the 7-day PAR and other questionnaires may lack the ability to capture activities lasting less than 10 minutes, or requiring less exertion than brisk walking. Conversely, Innerd and colleagues (2015) found significant associations between self-reported physical activity levels and objective accelerometer-based measures of sedentary behaviour, although their analysis is limited to adults aged 85 or older, where self-reported levels tend to be lower (Meijer et al., 2001), potentially bypassing the overestimation of vigorous levels described by Prince and colleagues (2008). Nonetheless, Ogonowska-Slodownik et al. (2022) stratified their analysis by age groups, and found that subjective and objective measures of physical fitness differed significantly across all age groups (60–65 years, 66–70 years, > 70 years). In addition, a trend for over-estimation of self-report moderate to vigorous physical activity was found (Ogonowska-Slodownik et al., 2022). In light of highly heterogeneous result patterns and poor overall consistency between measures, the general agreement in research appears to be that a combination of self-reported and objective measures offers a more holistic picture of physical capability. Therefore, adhering to consistent guidelines recommending the combination of different methods would be the optimal strategy to comprehensively assess physical fitness (Kowalski et al., 2012).

The largest contribution to the investigation of the agreement between self-report and objective physical fitness data comes from the systematic review performed by Prince and colleagues (2008), which generated inconsistent results as discussed above. Here, we will contribute towards the investigation of this association, by performing a secondary data analysis on data that was gathered from three cross-sectional studies conducted between 2017 and 2019, that will be further described below.

Methodology of the studies

A secondary data analysis was conducted on data that was pooled from 3 studies conducted between 2017 and 2019. The primary aim of these studies was to look at the cross-sectional correlation between physical fitness and cognitive functioning in adults aged 50 and older, based on the hypothesis that greater levels of physical fitness seem to be connected with optimal cognitive functioning. This was done using both self-report and performance-based measures to assess physical fitness. All three studies used the Physical Fitness Questionnaire (PFQ) to investigate self-report physical activity. The PFQ is a 15-item self-report questionnaire which measures strength, balance, and aerobic conditioning, and it is devised to evaluate the risk for dementia in the older population (Hollamby, et al., 2017) (See Appendix A for PFQ). High levels of physical fitness as measured with PFQ are correlated with lower cognitive deterioration in individuals with dementia (Hollamby et al., 2017). The PFQ is investigated here as it affords a specific insight into questionnaire development that improves test validity.

As discussed, the three original studies primarily looked at the cross-sectional association between physical and cognitive test performance in older adults. The secondary aim of these studies was investigating how individual differences in cognitive reserve as measured through the Cognitive Reserve Index questionnaire (CRIq) were related to cognitive levels in older adults. Study 1 and 3 also assessed olfactory function through the Olfactory Function Field Exam (OFFE), and investigated its correlation to cognitive function in older adults. State and Trait anxiety were also measured as a control variable for the correlational analysis between physical and cognitive functioning. Although these measures are beyond the scope of our investigation, a summary of all the measures used in each study can be found in Table 5 below. I personally collected the dataset from Study 3 as part of my MSc dissertation, in collaboration with another MSc student from Birkbeck, University of London.

Table 5: Summary of the measures employed in the studies that were used for secondary data analysis

Study 1	**Study 2**	**Study 3**
- Cognitive Reserve Index questionnaire (CRIq) - Physical Fitness Questionnaire (PFQ) - Grip strength, pulse, SpO2, balance - Montreal Cognitive Assessment (MoCA) - Olfactory Function Field Exam (OFFE) - Prospective and Retrospective Memory Questionnaire (PRMQ)	- Cognitive Reserve Index questionnaire (CRIq) - Physical Fitness Questionnaire (PFQ) - Grip strength, pulse, SpO2, chair rise test - Montreal Cognitive Assessment (MoCA) - State-Trait Anxiety Inventory (STAI)	- Cognitive Reserve Index questionnaire (CRIq) - Physical Fitness Questionnaire (PFQ) - Grip strength, pulse, SpO2, chair rise test, balance - Montreal Cognitive Assessment (MoCA) - Olfactory Function Field Exam (OFFE) - Prospective and Retrospective Memory Questionnaire (PRMQ) - State-Trait Anxiety Inventory (STAI)

Materials

Study 1 was conducted at University College London (UCL), whereas study 2 and 3 were conducted at Birkbeck, University of London. Global cognition was assessed using the MoCA across the three studies. PFQ scores ranged 15-75, with lower scores being associated with greater physical fitness levels in study 2 and 3. In study 1, the PFQ was reverse-scored, so that the higher the score the higher the levels of physical fitness. In order to have a consistent set of data, we reversed the scores from study 1. As for objective physical function, study 1 used measures of handgrip strength and balance tests, study 2 used handgrip strength and chair-stands test, whilst study 3 used handgrip strength, chair-stands test, and balance tests. All studies monitored resting heart rate (pulse).

The handgrip strength test measured upper limb strength using a CAMRY-EH101 electronic hand dynamometer, which provides an accurate measure of maximum handgrip strength: participants were asked to perform the strongest possible grip on the dynamometer handle, twice with each hand. The highest value obtained was used in the analysis. Greater scores were indicative of greater upper limb strength.

The chair stand test assessed lower limb strength and endurance. This test was performed according to the procedure standardised by Jones and colleagues (1999) and it measured the number of stands that participants could complete with their arms folded across their chest in 30 seconds. A higher number of stands was indicative of greater lower limb strength and endurance.

The Stork balance test assessed the ability to balance on one foot. The participant was asked to place both hands on their hips, and one foot was raised and was placed against the inside of the knee of the supporting leg. A stopwatch was started and participants were asked to balance on the supporting leg for as long as possible. The stopwatch was stopped if the hands came off the hips, the supporting foot moved in any direction or the non-supporting foot lost contact with the knee. The final score was the total time in seconds that the participant could stay on one foot in the best out of three attempts. Greater balance times were indicative of greater overall fitness.

Finally, pulse was measured using a Contec - CMS50D finger pulse oximeter. Participants were asked to sit still for 5 minutes as the pulse and oxygen saturation were recorded every minute. The average of the last 3 recordings was used for the analysis. Lower pulse values were associated with greater overall fitness.

Participant demographics

The three studies followed the same inclusion criteria: participants needed to have normal or corrected-to-normal vision and hearing, English as first language, no current psychiatric disorders, no history of neurological disorders, and lastly no cardiovascular disease, due to the physically demanding nature of some of the objective physical measures. No participants were identified as extreme outliers during original data collection, therefore no data cleaning was performed. For the purpose of the current investigation, eight participants were excluded from the analysis due to missing data on the PFQ, which was essential to perform our analysis. This resulted in a final sample size of N= 271 (188 females,83 males) of older adults aged 50 and over, with ages ranging from 50 to 91, M = 69.52, SD = 8.41. Study 1 had 111 participants (75 females, 36 males), aged 50 or over, with ages ranging from 50 to 91, M = 66.90, SD = 11.00. Study 2 consisted of 93 participants (64 females, 29 males) over the age of 60, with ages ranging from 60 to 87, M = 71.23, SD = 5.15. Study 3 involved a total of 67 adults (49 females, 18 males) over the age of 55 years old. The age of participants ranged from 55 to 91, M = 71.49, SD = 5.72. The three studies included a total of 271 participants aged 50 and over.

Data analysis

The agreement between self-reported and objective measures of physical fitness was analysed using correlation analysis. Shapiro-Wilk tests for normality showed that the data for PFQ, grip strength, balance, and chair-stands test were not normally distributed ($p > .05$). Therefore, Spearmans's correlation coefficient was used to establish the magnitude of the correlation between the variables under investigation, with the level of significance set at .05. Evidence shows that physical activity levels tend to significantly differ by age among older adults, with older age groups (80-85) showing as much as 50% lower physical activity levels than younger adults (Lohne-Seiler et al., 2014). Our analysis looked at the influence of age on the association between self-reported and objective physical fitness, by conducting partial correlation analyses controlling for age, and comparing these to zero-order correlations. Finally, stepwise multiple regression analyses were conducted, where subjective and objective measures of physical activity predicted the cognitive outcome. Missing values were excluded from the analyses. All analyses were performed using SPSS version 26.

As described above, the studies employed different measures of objective physical function: study 1 used handgrip strength and balance tests, study 2 used handgrip strength and chair-stands test, whilst study 3 used handgrip strength, chair-stands test, and balance tests, and all studies monitored heart rate. The analyses focused on correlating different combinations of measures based on data availability, in order to maximise sample sizes where possible. Handgrip strength and pulse were the only objective variables available for the overall dataset composed of 3 studies. Both overall PFQ scores and PFQ strength scores were compared with the handgrip strength score for the complete dataset composed of 3 cohorts. Pulse was also compared to both overall PFQ scores and PFQ aerobic sub-scores for the complete dataset. For data from study 1 and 3, which both included an objective assessment of balance, balance measures were compared with the PFQ balance sub-score, as well as with the overall PFQ score, and a composite measure of objective physical fitness was created by averaging standardised scores for handgrip strength and balance measures, and compared to overall PFQ scores. The use of standardises scores, which were obtained by subtracting the sample mean from raw scores and dividing this number by the sample standard deviation, ensured that the composite score was derived from comparable units of measurement. For data from studies 2 and 3, chair stands were correlated to the strength sub-score of the PFQ, and the overall PFQ, and a composite measure of objective physical fitness was created by averaging standardised scores handgrip strength and chair-stands measures, and correlated to overall PFQ scores. For study 3, a composite objective physical activity measure was computed averaging standardised scores from the handgrip strength, chair-stands test, and balance tests. This was correlated with the overall PFQ scores. Finally, stepwise multiple regression analyses were conducted, where age, grip strength (as the only performance-based measure available for the whole dataset) and PFQ scores (overall scores, as well as individual components separately) were tested as predictors of MoCA scores. Grip and PFQ were entered in opposite order in two different analyses, so as to test their unique contribution in explaining the variance in the MoCA. Further stepwise regression models were conducted, where the three subcomponents of the PFQ were entered as separate predictors after age and before grip, in order to assess whether grip explained any additional variance that was not already accounted for by age and the PFQ subcomponent.

Results

Table 6 below presents descriptive statistics divided by each study and for the overall sample.

Table 6: Descriptive statistics divided by study and for the overall sample.

	Age		Gender		Years of Education		PFQ Aerobic		PFQ Strength		PFQ Balance		PFQ Overall		Handgrip		Balance		30 seconds		MoCA		SpO2		Pulse	
	M	**SD**	**Female**	**Male**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Study 1 (N= 111)	66.9	11.0	75	36	12.1	5.2	13.8	5.4	11.5	5.1	14.8	7.0	40.1	15.2	21	11	1.7	2	-	-	25.3	4.9	95.4	2	71.4	2.3
Study 2 (N= 93)	71.2	5.2	64	29	16.7	4.1	8.8	3.8	8.2	2.9	8.3	3.7	25.3	7.9	26.9	9.1	-	-	14.4	5.2	27.3	2.3	95.6	3	67.9	9.5
Study 3 (N= 67)	71.5	5.7	49	18	16.3	3.0	9.1	3.8	8.3	2.9	8.9	3.9	26.2	8.4	26.1	7.8	37.4	25	14.2	5.1	28.9	1.5	95.1	5	74.3	11.3
Total (N= 271)	69.5	8.4	188	83	15.0	1.1	10.9	5.2	9.6	4.3	11.1	6.2	31.6	13.5	24.3	9.9	15.2	23	14.3	5.2	27.7	3.7	95.4	4	70.9	10.3

PFQ = Physical Fitness Questionnaire, MoCA = Montreal Cognitive Assessment, SpO2 = Blood oxygen saturation. M = mean, SD = Standard Deviation.

3-study analysis

When considering the overall dataset obtained by merging the three studies, a highly significant decline in physical fitness is observed in relation to age, both as measured through handgrip strength, ρ = -.267, N = 271, p <.001, and PFQ, ρ = .122, N = 271, p =.044 (greater PFQ scores indicate lower self-reported physical fitness levels).

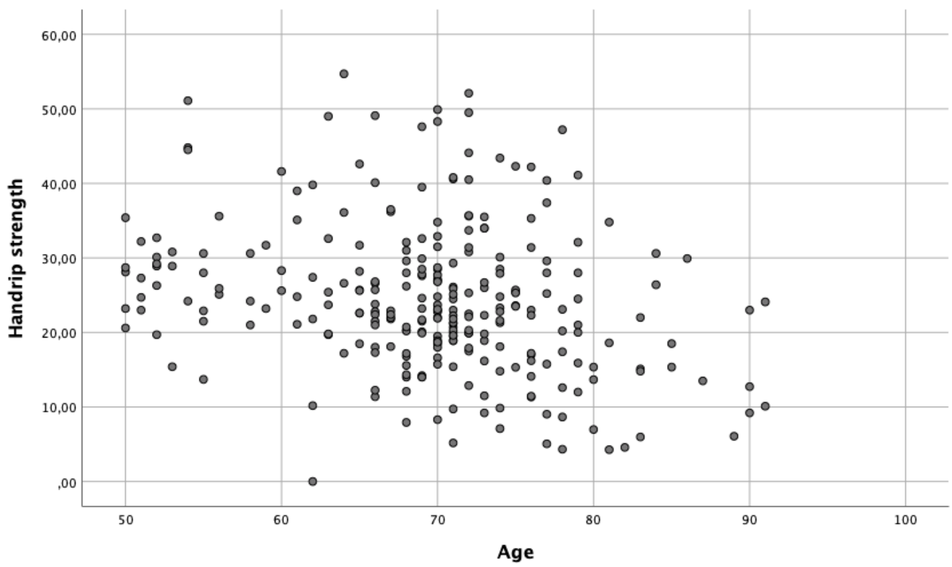

Figure 19: Scatterplot showing the correlation between age and handgrip strength

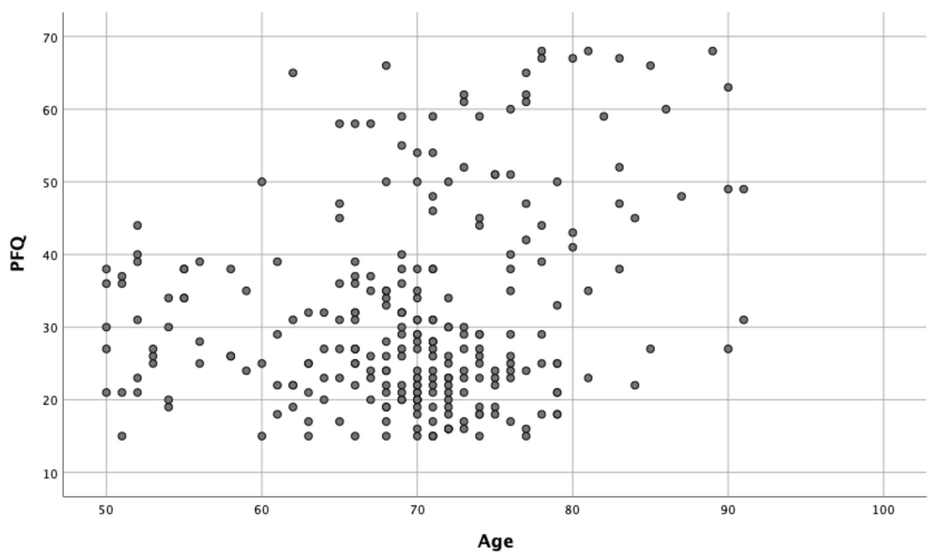

Figure 20: Scatterplot showing the correlation between age and PFQ

Handgrip strength was the only performance-based variable available for the three cohorts. Therefore, we compared overall PFQ scores and PFQ strength sub-scores with the objective handgrip strength score for the complete dataset composed of data merged from all 3 studies. 8 participants had missing

PFQ data, therefore they were excluded from the analysis, resulting in a final sample size of N = 271. The possible scores were divided into equal intervals, and the sample was also categorised into high (PFQ ranging 15-35), medium (PFQ ranging 35-55) or low (PFQ ranging 55-75) self-report physical fitness groups, according to their self-reported physical fitness levels. 68% of participants (N= 184) reported high levels of physical fitness, 23% reported intermediate levels of physical fitness (N = 62), and 9% (N= 25) reported low levels of physical fitness. The mean reported PFQ was M = 31.59, SD = 13.53, while the mean handgrip strength for the sample was M = 24, 43 kg, SD = 9.81 kg.

The analysis showed a significant negative correlation between overall PFQ scores and handgrip strength, $\rho = -.517$, N = 270, $p < .001$, indicating that higher handgrip strength scores are correlated with lower PFQ scores, which in turn indicate greater self-report physical fitness levels. This correlation was still highly significant when running a partial correlation analysis controlling for the effects of age, $r = -.508$, df = 267, $p < .001$.

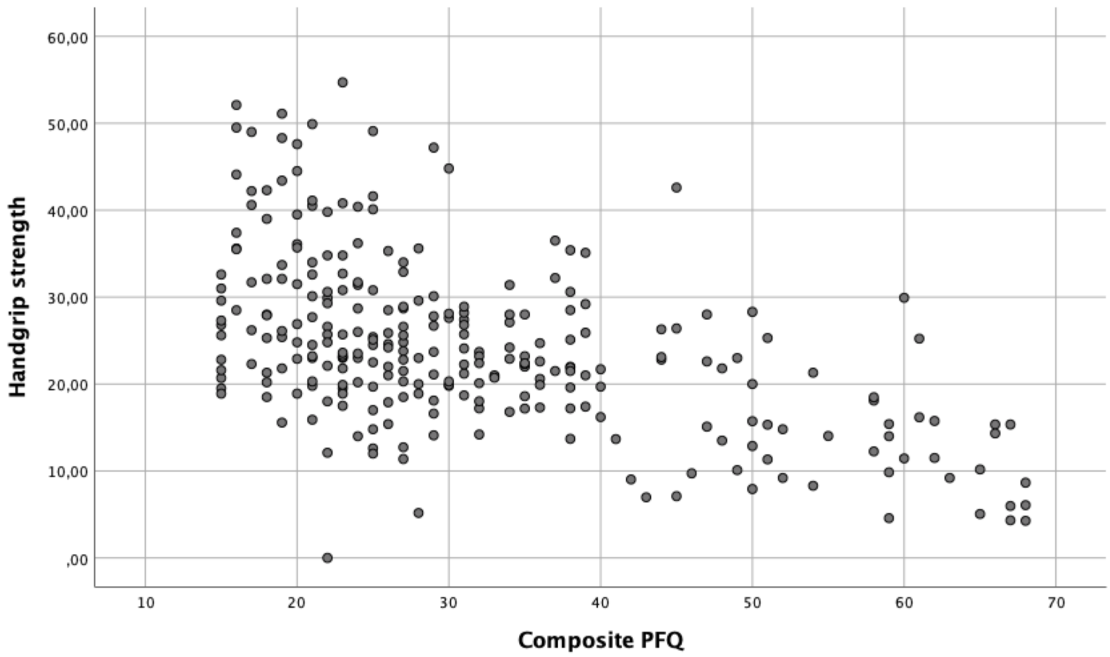

Figure 21: Scatterplot for the correlation between PFQ scores and handgrip strength

The correlation between overall PFQ and handgrip strength was still present, albeit weaker in each group, when comparing low, medium and high self-report PFQ and handgrip strength levels (see table 7 and figures 22-24 below).

Table 7: Coefficients for the correlations between PFQ and grip strength divided by PFQ groups

	ρ	p	N
High PF	-.31	<.001	183
Medium PF	-.35	.005	62
Low PF	-.49	.013	25

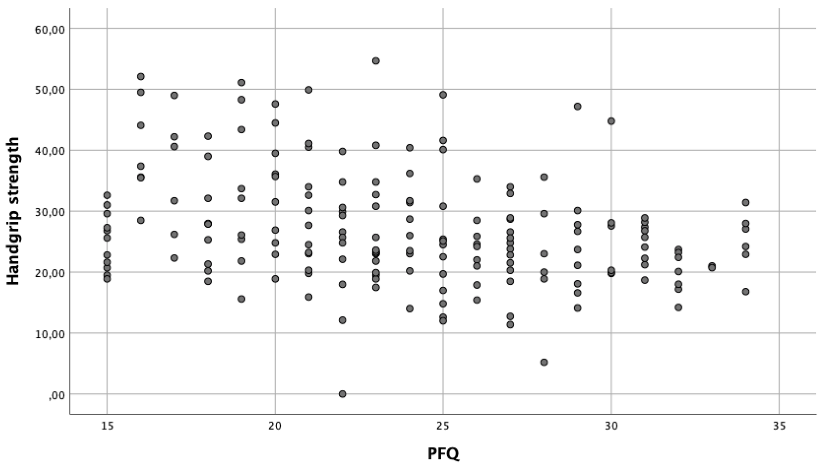

Figure 22: Scatterplot showing the association between PFQ and handgrip strength in the higher PF group

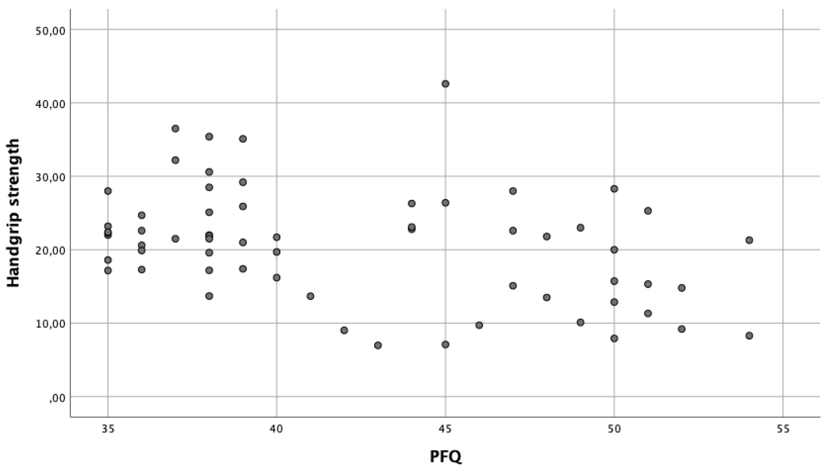

Figure 23: Scatterplot showing the association between PFQ and handgrip strength in the intermediate PF group

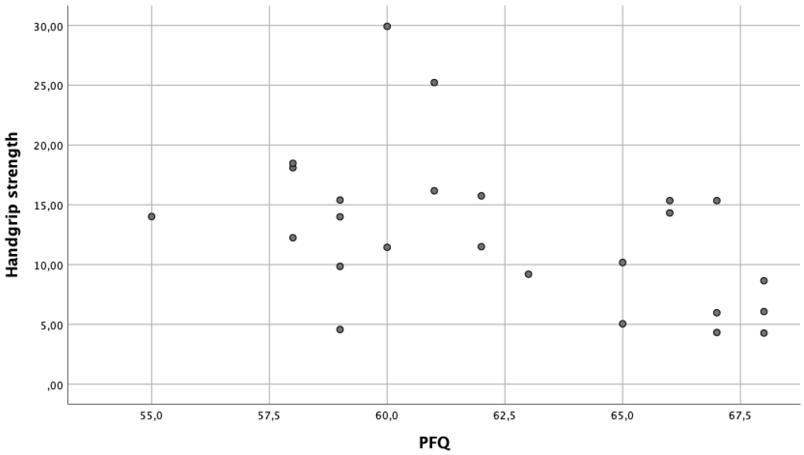

Figure 24: Scatterplot showing the association between PFQ and handgrip strength in the low PF group

The correlation between overall PFQ and handgrip strength was still significant when stratifying the analysis by younger (50-64), medium (65-79), and older (80 to 91) age groups (see table 8 below). Although the correlation only approached significance in the younger age group, this may also be related to the sample size being smaller to allow for a group division by age, or to younger participants showing a tendency to overestimate their activity levels (See Figure 25 below).

Table 8: Coefficients for the correlations between PFQ and grip strength divided by age groups

	ρ	p	N
Younger (50-64)	-.26	.050	56
Medium (65-79)	-.53	<.001	191
Older (80-92)	-.57	.004	23

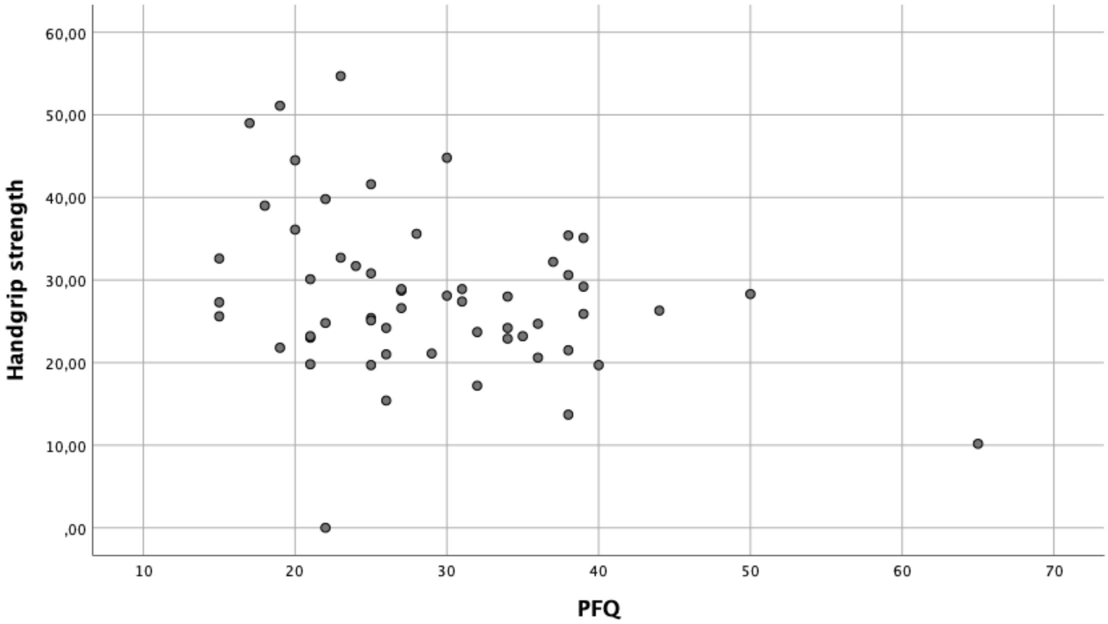

Figure 25: Scatterplot showing the association between PFQ and handgrip strength in the younger age group

The association between objective and self-report measures of strength was confirmed by further analyses showing a significant, moderate, negative correlation between the PFQ strength sub-scores and handgrip strength, $\rho = -.543$, $N = 270$, $p < .001$, where the magnitude of the association was greater than that observed with composite PFQ scores. A partial correlation analysis controlling for age suggested that these results were not driven by age effects, $r = -.508$, $df = 267$, $p < .001$.

The analysis looking at the correlation between pulse and overall PFQ showed a significant, but weak correlation between overall PFQ and pulse, $\rho = .188$, $N = 270$, $p = .002$, which continued to be significant when controlling for age in a partial correlation analysis, $r = .175$, $df = 267$, $p = .004$. This correlation persisted when looking at the association between the aerobic conditioning scores of the PFQ and pulse, $\rho = .192$, $N = 270$, $p = .001$.

Analysis on Studies 1 and 3

A composite performance-based physical fitness score was computed using standardised scores for balance and handgrip strength scores from studies 1 and 3. This score showed a significant negative correlation with overall PFQ scores, ρ = -.662, N = 177, p <.001. This correlation was still observable when controlling for the effects of age, r = -.627, df = 174, p < .001.

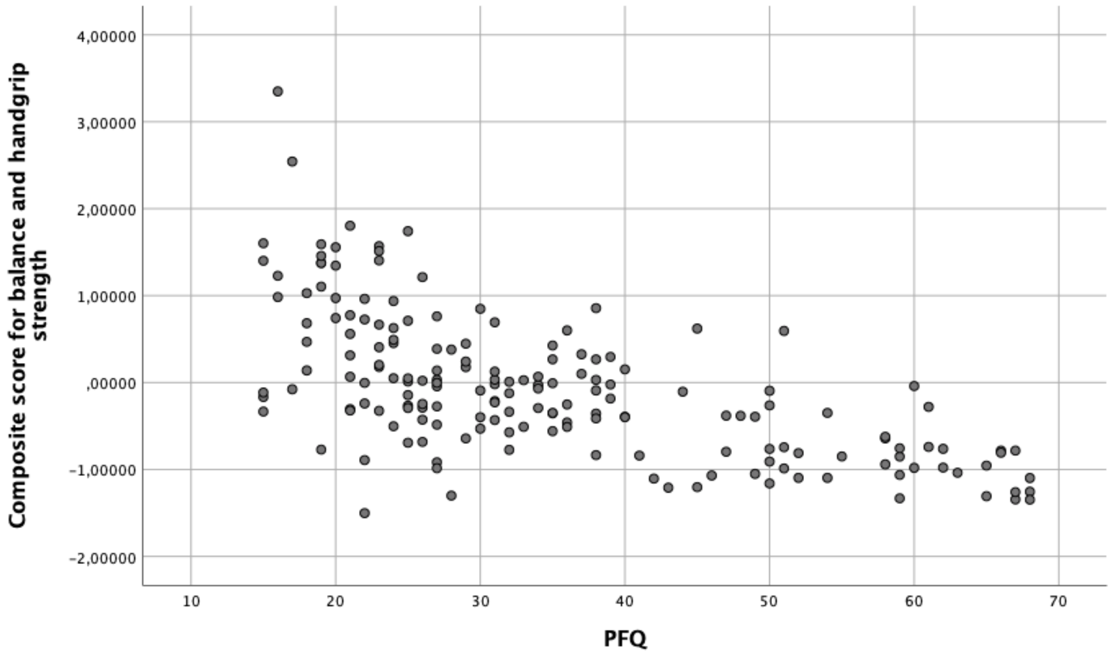

Figure 26: Scatterplot showing the correlation between a composite performance-based measure derived by averaging balance and handgrip strength measures, and composite PFQ scores

Performance in the balance test was significantly correlated with the PFQ balance sub-score, ρ = -.617, N = 177, p <.001, and with the overall PFQ score, ρ = -.622, N = 177, p <.001. These correlations were still observable when controlling for the effects of age, both for the balance sub-score of the PFQ, r = -.553, df = 174, p < .001, and for the overall PFQ, r = -.527, df = 174, p < .001

Analysis on Studies 2 and 3

A composite measure of objective physical fitness was created by using standardised scores for handgrip strength and chair-stands measures. This score showed a significant negative correlation with overall PFQ scores, ρ = -.533, N = 160, $p <.001$. This correlation was still observable when controlling for the effects of age, r = -.498, df = 157, $p < .001$.

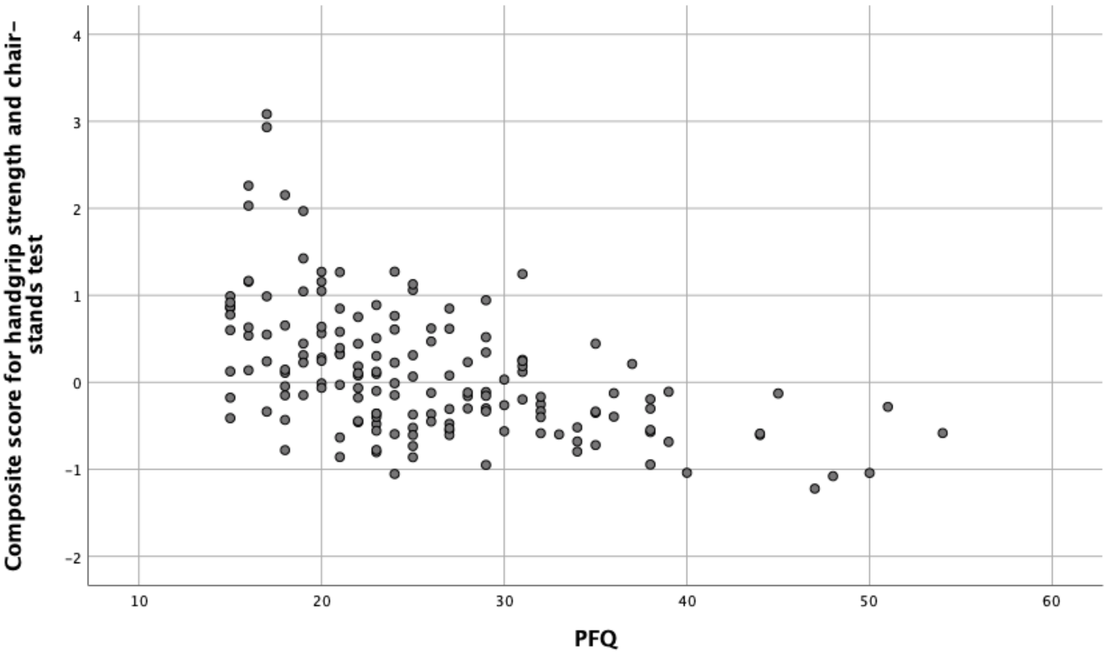

Figure 27: Scatterplot showing the correlation between a composite performance-based measure derived by averaging chair stands tests and handgrip strength measures, and composite PFQ scores

Chair-stand test scores were significantly correlated with the PFQ strength sub-score, ρ = -.240, N = 160, p = .002, and with the overall PFQ score, ρ = -.417, N = 160, $p <.001$. These correlations were still observable when controlling for the effects of age, both for the strength sub-score of the PFQ, r = -.188, df = 157, p = .018, and for the overall PFQ, r = -.426, df = 157, $p < .001$

Analysis on Study 3

A composite objective physical fitness measure was computed by using standardised scores from the handgrip strength, chair-stands test, and balance tests. This measure was significantly correlated with overall PFQ scores, ρ = -.599, N = 67, p <.001, and this association was still significant when controlling for age in a partial correlation analysis, r = -.526, df = 64, p < .001.

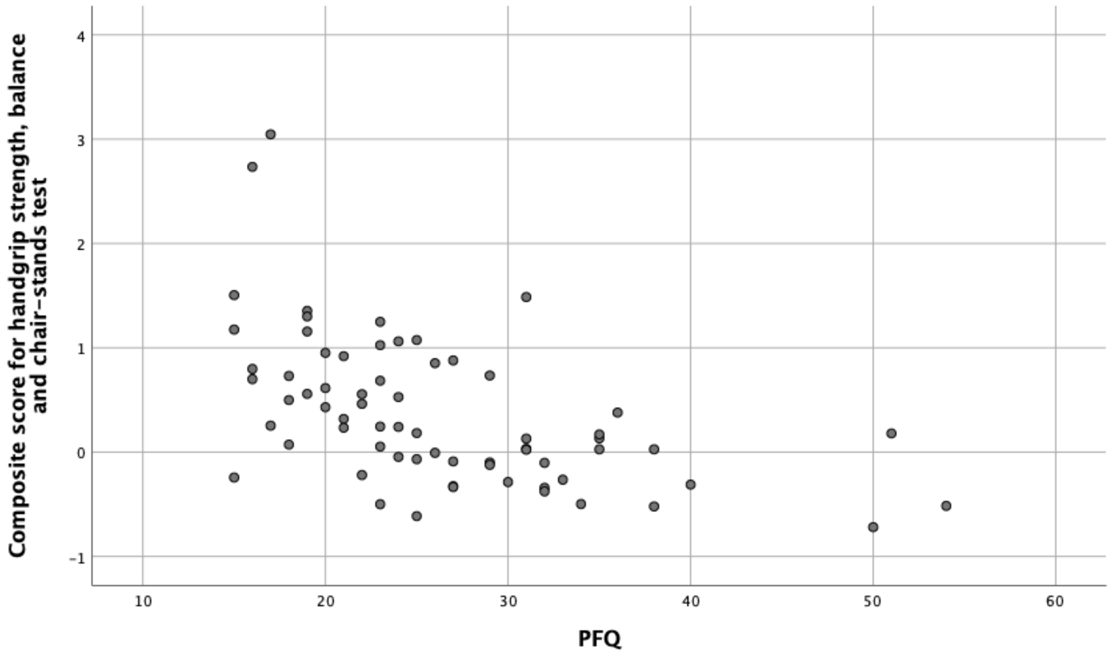

Figure 28: Scatterplot showing the correlation between a composite performance-based measure derived by averaging balance, chair stands and handgrip strength tests measures, and composite PFQ scores

Regression analyses

In a stepwise multiple regression analysis, age was entered as the first predictor, followed by PFQ and grip strength. Age was entered first, and accounted for 10% of the variance in MoCA, R^2 =10%, SE = .02, F(1, 268) = 30.90, p < .001. The PFQ was entered on the second step, and accounted for a further 15% of the variance in MoCA, R^2= 25%, SE= .02, F(2,267) = 44.14, p < .001. Age shared 6% of variance predicted by the PFQ. Age predicted a unique variance of 4% in MoCa scores. Grip strength was entered on the third step, and did not account for any further variance in MoCA, R^2= 25%, SE= .02, F(3,266) = 29.57, p < .001. Age and PFQ shared 15% of variance predicted by grip strength. Results of the stepwise regression are presented in Table 9 below.

Table 9: Regression coefficients for the stepwise regression with Age, PFQ and Grip onto MoCA.

Predictor variable	R	R^2	R^2 change	Part correlation (squared)	B	Standard error	Beta	t
Age	.32	.10	.10***	$-.21^2$= .04	-.09	.02	-.22	-4.00***
PFQ	.50	.25***	.15***	$-.31^2$= .10	-.10	.02	-.37	-5.79***
Grip	.50	.25***	.00	$.04^2$=.00	.02	.02	-.05	.76

& *p* < .10 * p < .05; ** p < .01; *** p < .001

A further stepwise regression analysis was conducted, where age was entered as the first predictor, followed by grip strength and PFQ. Age accounted for 10% of the variance in MoCA, R^2 =10%, SE = .02, F(1, 268) = 30.90, p < .001. Grip strength was entered on the second step, and accounted for a further 5% of the variance in MoCA, R^2= 15%, SE= .02, F(2,267) = 24.58, p < .001. Age shared 6% of variance predicted by grip strength. PFQ was entered on the third step, and accounted for further 9% of variance in MoCA, R^2= 25%, SE= .02, F(3,266) = 29.57, p < .001. Age and grip shared 15% of variance predicted by PFQ. Results of the stepwise regression are presented in Table 10 below.

Table 10: Regression coefficients for the stepwise regression with Age, Grip and PFQ onto MoCA.

Predictor variable	R	R^2	R^2 change	Part correlation (squared)	B	Standard error	Beta	t
Age	.32	.10	.10***	$-.21^2 = .04$	-.09	.02	-.22	-4.00***
Grip	.39	.15***	.05***	$.04^2 = .00$	.02	.02	-05	.76
PFQ	.50	.25***	.09***	$-.31^2 = .10$	-.10	.02	-.37	-5.79***

[&] *p* < .10 * p < .05; ** p < .01; *** p < .001

Further stepwise regression analyses were conducted, where the subcomponents of the PFQ (aerobic conditioning, strength and balance) were entered separately in each model after age and before grip, in order to assess whether grip strength explained any of the remaining variance in MoCA scores after age and the different components of the PFQ were entered.

In the third regression analysis, PFQ strength was entered as the third predictor after age and grip. Age accounted for 10% of the variance in MoCA, R^2 =10%, SE = .03, F(1, 268) = 30.90, p < .001. PFQ strength was entered on the second step, and accounted for a further 10% of the variance in MoCA, R^2= 20%, SE= .06, F(2,267) = 33.15, p < .001. Age shared 5% of variance predicted by PFQ strength. Grip was entered on the third step, and accounted for a further 1% in MoCA, R^2= 21%, SE= .02, F(3,266) = 23.01, p < .001. Age and PFQ strength shared 15% of variance predicted by grip. Results of the stepwise regression are presented in Table 11 below.

Table 11: Regression coefficients for the stepwise regression with Age, PFQ strength and Grip onto MoCA.

Predictor variable	R	R^2	R^2 change	Part correlation (squared)	B	Standard error	Beta	t
Age	.32	.10	.10***	$-.23^2 = .05$	-.11	.03	-.24	-4.23***
PFQ strength	.45	.20***	.10***	$-.23^2 = .05$	-.23	.06	-.27	-4.11***
Grip	.46	.21***	.01	$.08^2 = .01$	.04	.02	.10	1.54

[&] *p* < .10 * p < .05; ** p < .01; *** p < .001

In a further stepwise regression analysis, MoCA scores were predicted from age, PFQ aerobic conditioning and grip. Age accounted for 10% of the variance in MoCA, R^2 =10%, SE = .03, F(1, 268) = 30.90, p < .001. PFQ aerobic conditioning was entered on the second step, and accounted for a further 8% of the variance in MoCA, R^2= 19%, SE= .04, F(2,267) = 30.72, p < .001. Age shared 5% of variance predicted by PFQ aerobic conditioning. Grip was entered on the third step, and accounted for further 2% of variance in MoCA, R^2= 21%, SE= .02, F(3,266) = 23.35, p < .001. Age and PFQ aerobic conditioning shared 14% of variance predicted by grip. Results of the stepwise regression are presented in Table 12 below.

Table 12: Regression coefficients for the stepwise regression with Age, PFQ aerobic conditioning and Grip onto MoCA.

Predictor variable	R	R^2	R^2 change	Part correlation (squared)	B	Standard error	Beta	t
Age	.32	.10	.10***	$-.25^2$= .06	-.11	.03	-.26	-4.59***
PFQ aerobic	.43	.19***	.08***	$-.23^2$= .05	-.17	.04	-.24	-4.22***
Grip	.46	.21***	.02**	$.15^2$=.02	.06	.02	.16	2.68**

[&] *p* < .10 * p < .05; ** p < .01; *** p < .001

In a further stepwise regression analysis, MoCA scores were predicted from age, PFQ balance and grip. Age accounted for 10% of the variance in MoCA, R^2 =10%, SE = .03, F(1, 268) = 30.90, p < .001. PFQ balance was entered on the second step, and accounted for a further 15% of the variance in MoCA, R^2= 25%, SE= .04, F(2,267) = 44.07, p < .001. Age shared 7% of variance predicted by PFQ balance. Grip was entered on the third step, and did not account for any further variance in MoCA, R^2= 25%, SE= .02, F(3,266) = 29.51, p < .001. Age and PFQ balance shared 15% of variance predicted by grip. Results of the stepwise regression are presented in Table 13 below.

Table 13: Regression coefficients for the stepwise regression with Age, PFQ balance and grip onto MoCA.

Predictor variable	R	R^2	R^2 change	Part correlation (squared)	B	Standard error	Beta	t
Age	.32	.10	.10***	$-.18^2$= .03	-.09	.03	-.19	-3.44**
PFQ balance	.50	.25***	.15***	$-.31^2$= .10	-.23	.04	-.38	-5.80***
Grip	.50	.25***	.00	$.04^2$=.00	.02	.02	.05	.74

[&] *p* < .10 * p < .05; ** p < .01; *** p < .001

Discussion

The present study looked at the agreement between performance-based and self-reported measures of physical functioning, through the secondary analysis of cross-sectional data from 3 studies including both objective and self-reported measurements. Overall, our findings indicated low to moderate rates of agreement between self-reported and objective measures of physical function (Spearman's ρ ranging from -.66 to -.17). Conversely, the largest study to date (Prince et al., 2008) found inconsistent rates of agreement, the absolute values of which ranged from .71 to .96, with no clear trends of association emerging from the results, and self-reported data being both higher and lower than the objective measures. Findings from our study are broadly in agreement with further studies looking at the convergence between physical activity questionnaires and accelerometer-based data, which have reported criterion validity of ρ = .030 (Craig et al., 2003) for greater sample sizes (N = 2721). Our results are also in line with additional findings from a recent analysis conducted by Innerd and colleagues (2015), which yielded Spearman's ρ ranging from .10 to .52, (N = 337), suggesting a modest agreement between self-report physical activity and accelerometer data. Although our study did not employ accelerometer data, the performance-based measures that we utilised showed similar values of convergent validity with self-report measures. Our study also showed a significant decline of physical fitness with age, both as assessed with PFQ and performance-based measures, which is in agreement with previous studies showing an age-related decline of physical functioning (Lohne-Seiler et al., 2014; Ogonowska- Slodownik et al., 2022). Finally, our results showed that in a regression model predicting cognitive scores based on age, PFQ and grip strength sequentially, grip did not explain any additional variance that was not already accounted for by the PFQ. These results point towards a convergence between self-reported and performance-based measures of physical fitness, and are highly relevant for scale validation of the PFQ, as they indicate that the PFQ and grip are assessing the same construct. Indeed, the fact that grip does not explain any further variance in cognitive scores than the one already accounted for by the PFQ suggests that grip strength and the PFQ are measuring the same thing, because the variance in cognitive scores explained by grip strength overlaps with the variance in cognitive scores already explained by the PFQ.

Prince et al. (2008) found the output generated from self-reported data to be both higher and lower than that obtained from performance-based measures, with no clear patterns in the data analysed. The authors highlighted the consequent need for more reliable assessment protocols to measure physical fitness. Specifically, the authors highlighted the predominant use of accelerometer-based measures to objectively assess physical functioning, which do not capture activities involving the upper limbs.

In contrast, our analyses used a comprehensive assessment of physical function, including both upper and lower limb functioning. Our results consistently generated significant, negative correlation coefficients, which indicated that greater performance in objective measures of physical fitness was correlated with lower PFQ scores (where lower PFQ scores indicate greater self-reported physical fitness levels). A negative correlation was found for all measures: both handgrip strength and pulse measures across all studies, as well as a composite measure of balance and handgrip from studies 1 and 3, a composite measure of chair-stands and handgrip from studies 2 and 3, and a measure derived from handgrip, balance, and chair-stands from study 3, all showed a significant negative correlation with PFQ. This was also the case when looking individually at balance, pulse, handgrip strength and chair-stands in correlation with the PFQ sub-scores of balance, aerobic conditioning, and strength respectively.

Where data from all three studies were analysed, the analysis was stratified by younger (50-64), medium (65-79), and older (80 to 91) age groups. When comparing these groups, the correlation between overall PFQ and handgrip strength was still significant in each group. As discussed in the results section, the magnitude of the correlation was lower and only approached significance in the younger age group. While it can be argued that this may be related to the sample size being reduced in each group when dividing the sample by age categories, there was also a slight trend for overestimation of self-reported physical fitness levels in the younger age group. In fact, there is a tendency for younger adults to report greater levels of activity than their older counterparts (Zimmermann-Sloutskis et al., 2010), and this may be reflected in our sample. Similarly, when dividing the 3-study analysis by self-reported physical fitness levels (low, medium, high), the correlation between handgrip strength and PFQ was attenuated in the high and medium self-reported groups compared to when considering the data in its entirety, while the magnitude of the correlation remained consistent in the low self-reported physical fitness group. When inspecting figures illustrating this correlation by physical fitness groups (figures 22 to 24), a certain degree of over and underestimation of physical functioning emerges in the intermediate and higher self-reported physical fitness groups respectively, which may explain why the correlation was attenuated in these groups.

Convergent validity

The concept of convergent validity may be of use in order to contextualise our results and interpret the magnitude of the correlations that emerged from our analysis. Convergent validity is usually investigated when validating questionnaires, and it verifies whether the output of a questionnaire is consistent with related measures (Abma et al., 2016). In other words, convergent validity tests whether the scores of any given measure correlate with scores of other related measures assessing the same construct. Convergent validity is usually deemed to be adequate when the correlation with a measure assessing the same construct is at least at the .50 level (Abma et al., 2016). Although our analysis did not have the objective of validating a new measure, the guidelines for adequate convergent validity are useful to interpret the magnitude of the correlation coefficients yielded by our analysis, which ranged from -.66 to -.17. Considering our results in light of this acceptability threshold, the PFQ seems to be a fairly valid measure for self-reported physical fitness when compared to an objective handgrip strength measure, although the correlation coefficient of -.52 suggests only a modest adequacy of the PFQ. On the other hand, the magnitude of the correlation between the PFQ and resting heart rate (-.19) seems to point against a convergence between these measures. It should be considered that within our sample, there was a high prevalence of participants taking heart medication (although this variable was not formally recorded), which may have influenced the resting heart rate values and deflated the magnitude of this correlation. Further to this, handgrip strength provides an index of performance, as opposed to informing on a resting state measure, therefore it may encompass physical function more accurately.

As far as the individual sub-scores of the PFQ and the composite performance-based scores that we looked at, both the composite measure derived from balance and handgrip strength, the measure derived from chair-stands and handgrip strength, and the measure derived from balance, handgrip strength and chair-stands were above the appropriate convergent validity threshold when correlated with the PFQ, although once again the correlation coefficients of -.66, -.53 and -.60 respectively indicate only a modest convergent validity of the PFQ. When considered individually against the balance sub-score of the PFQ, objective balance measures showed moderate correlation ($\rho = -.62$), whilst the chair-stands test did not meet the adequacy threshold of .50 when compared to the strength sub-score of the PFQ, with a correlation coefficient of -.24. However, this may be related to the chair-stands test not purely assessing strength, but also involving other components of physical fitness such as dynamic balance and general mobility, given that the validity increased up to -.42 when comparing performance in the chair-stands test with the overall PFQ scores.

PFQ and Grip strength: overlapping variance explained in MoCA scores

The regression analysis predicting MoCA scores based on age, PFQ and grip strength sequentially showed that grip did not explain any residual variance that was not accounted for by overall PFQ scores. This trend was confirmed by a further regression analysis predicting MoCA scores from age, the PFQ strength component, and grip sequentially: results from this analysis showed that grip did not explain any additional variance that was not already accounted for by PFQ strength. This finding suggests that grip and the strength component of the PFQ may be capturing the same construct, and is in line with the measurement properties of the PFQ strength subscale, which investigates exclusively aspects of grip strength, daily life tasks requiring physical strength, or lifting abilities. In other words, when added after either PFQ strength or PFQ overall scores, grip does not explain any further variance that was not already accounted for by age and PFQ. We propose that this is because grip strength and PFQ (both overall PFW and the PFQ strength sub-score) are measuring the same construct. Therefore, when grip is entered in the regression model after the PFQ, the variance in the cognitive outcome explained by grip overlaps with that already explained by the PFQ, because both the PFQ and grip strength are capturing the same construct. Therefore, it is reasonable to think that grip strength does not explain an additional portion of variance in the cognitive outcome, because all of the variance in MoCA that grip explains has already been accounted for by its convergent measure, the PFQ, which was already entered in the analysis. The pattern of findings was different when the same analysis was performed with the PFQ aerobic component as the second predictor: here, results showed that grip explained additional variance that was not already explained by the PFQ aerobic scores. This is coherent with the characteristics measured by the aerobic component of the PFQ, which investigates aspects of walking abilities and aerobic exertion that do not overlap with any aspect of muscular strength that is captured by a strength-related measure such as grip. A further regression predicting MoCA scores based on age, the PFQ balance component, and grip, showed that grip did not explain any additional variance not already accounted for by the balance component of the PFQ. Again, these results are in line with the measurement properties of the balance component of the PFQ, which enquires on the frequency at which respondents "lose balance while standing up or sitting down". While capturing balance abilities, this question undoubtedly involves aspects of lower limb strength that could justify the pattern of results that we observed, and motivate the absence of additional variance explained by grip. Conversely, when running a model that predicts MoCA scores based sequentially on age, grip and overall PFQ, our results showed that there was some residual variance explained by the PFQ, which is likely explained by the PFQ including questions on balance and aerobic conditioning, that are not likely to be captured by an objective measure such as grip, which strictly evaluates strength.

The results of these regression models are highly relevant. Firstly, they substantiate the construct validity of the PFQ, i.e. the extent to which this questionnaire accurately measures what it intends to measure. In fact, the pattern of results that we observed, whereby grip did not explain any additional variance that was not already accounted for by the PFQ, is only possible in the case that the PFQ and grip strength are assessing *the same construct.* Secondly, the patterns of overlapping variance that emerged from our analysis have great value for the development and validation of physical fitness questionnaires. Indeed, the analysis that we performed here could be implemented in scale development, following a process whereby single sub-components of a questionnaire are tested against an objective measure assessing the same construct, in order to investigate whether both measures explain any overlapping variance of a third variable that both measures are related to. Ideally, this process could be done in an iterative fashion, whereby sub-components can be modified until the variance explained by both measures reaches a perfect overlap. This process would also lend greater credibility to convergent validity analyses. In fact, findings from our regression analyses lend further support to the results from our correlational analyses, where sub-components of the PFQ were simply correlated with their corresponding objective measure, by demonstrating overlapping explained variance in a third variable that both subjective and objective measures are related to. This process could be extended to investigating the shared variance of single questionnaire items against their objective counterparts. Our study only looked at the grip component as it was the only measure available across the three studies, and therefore allowing for an extensive analysis. Further directions of research could involve the repetition of this analysis on the other sub-components of the PFQ, in larger datasets including an objective assessment of balance and aerobic conditioning that could be analysed against the related counterparts of the PFQ.

Overall, our results seem to point towards a modest convergence between the performance-based measures that we used and the PFQ and its sub-components. While significant correlations indicating a certain degree of agreement between objective and self-reported measures were observed, they only marginally met the criteria deemed as adequate convergent validity for related constructs. The modest correlation between self-report and objective measures may reflect issues with participants' recall and interpretation of self-report measures, or an only marginal ability of questionnaires to measure the construct of interest. The modest rate of agreement emerging from our analysis is especially concerning in light of the fact that these measures are employed interchangeably in epidemiological research. Given that findings based on self-report data are used to investigate potential risk factors leading to adverse health outcomes and to inform the development of interventions and public health policies, the use of a combination of performance-based and self-report measures to assess physical fitness should be encouraged. This would allow to further assess the validity of self-report measures, and to investigate risk factors for adverse outcomes with greater reliability. An iterative analysis using objective and self-reported measures of physical fitness, investigating overlapping explained variance of a third related construct is also suggested, as it lends further credibility to convergent validity analyses and maximally exploits the simultaneous use of subjective and objective measures of the same variable.

A strength of our study was the use of different performance-based measures of physical functioning. The majority of research looking at the convergent validity of self-reported measures uses accelerometer data, which is largely considered to be the gold standard method to validate self-report physical activity measures (Sirard & Pate, 2001). Nonetheless, the use of accelerometers to validate self-report measures of physical activity has been deemed as problematic, because accelerometers assess aspects of physical activity that may not reflect the aspects captured by self-reported measures (Leinonen et al., 2017), such as activities involving the upper limbs, or swimming. Our assessment of physical function included objective performance-based measures which reflected the individual components that the PFQ assessed, and allowed for a thorough investigation of convergent validity based on balance, aerobic conditioning, and strength measures.

Self-reported physical activity levels may be assessed using activity logs, diaries, or questionnaires. The 7-day PAR was the predominantly used questionnaire in Prince et al.'s (2008) systematic review of the literature. This questionnaire is a semi-structured interview, where participants are asked to recall time spent sleeping, working, or doing different physical activities in the past 7 days. Questions of the 7-day PAR include "Compared to your physical activity over the past 3 months, was last week's physical activity more, less, or about the same?". Question phrased as such introduce recall bias, as they enquire on activity levels over a relatively long period of time, and require participants to provide an estimated comparison of their activity levels over two different time periods. The otherwise most widely employed questionnaire in epidemiological research is the IPAQ, which was one of the most frequent questionnaires in the studies retrieved in Prince et al.'s (2008) review. Questions from the IPAQ enquire on the time spent being physically active on moderate or vigorous activities during the last 7 days. Aside from introducing recall bias, this type of questions may introduce the variable of the participant's perception of the intensity of physical activities. This perception has been shown to vary according to the physical status of the individual: for instance, individuals with higher body weight may perceive physical activities to be higher in intensity (Hagstromer et al., 2010). For example, the following question from the IPAQ: "During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, […]?" may introduce some uncertainty, as respondents may have different perceptions of what "heavy lifting" constitutes, based on their physical status. The problematic nature of some of the questions included in the 7-day PAR and the IPAQ, which were the predominant measures in Prince et al.'s review (2008), may explain the inconsistent pattern of results observed by the researchers.

The PFQ used in our study bypassed this issue as it did not involve participants' perception of the different intensities of physical activity. The PFQ involved three sections: one dedicated to aerobic conditioning, another to strength training, and a final one to balance and stability. Participants were asked to rate the frequency at which they encountered difficulties carrying out different activities within each of these dimensions. Overall, the activities described by the questionnaire appropriately capture physical functions that are normally carried out by older people (walking, climbing two flights of stairs, etc.). The nature of the questions, which enquire about difficulties in performing daily life activities, may also contribute towards minimising recall bias. Indeed, this type of question encompasses procedural aspects of daily life which may be less reliant on memory recollection (eg., "How often do you encounter difficulties in performing [physical activity]?), as opposed to questions enquiring on activity levels directly (e., "How many times have you exercised in the past month?"). Lastly, the PFQ was easily administered by researchers trained in its delivery, and well tolerated by participants.

This study had some limitations. Firstly, the sample size was relatively small, which made the stratification of the analysis by age or physical fitness groups problematic as the sample size was reduced even further. The three datasets that were considered included different combinations of performance-based measures, therefore the analysis on the complete dataset was only conducted for handgrip strength and pulse, which may have influenced the magnitude of the correlation coefficients that were observed. Further, the regression analysis looking for overlapping variance in MoCA scores could only be conducted for grip strength. Nonetheless, an attempt was made to maximise the sample size of each analysis and to create composite measures derived from performance-based components, based on data availability. While the use of different balance and strength performance-based measures has its strengths when compared to accelerometer-based data, as it may capture aspects of performance that the latter does not consider, the majority of existing research looking at convergent validity between self-report and objective measures has been conducted using accelerometer data. This study did not include accelerometer data, which would have been useful to assess the consistency with the alternative measures of physical performance that we used. Further, conducting multiple correlation and regression analyses without adjusting the significance threshold carries an increased risk of introducing type I errors, caused by the results reaching significance by chance, because the probability of finding at least one significant result due to chance increases as more tests are conducted. However, given the number and variety of hypotheses that out study aimed to address, the approach of multiple testing seemed to be the only suitable option. Finally, there was an issue with self-selection bias, whereby the older adults that were included in these studies had to be physically fit enough to reach the testing location, and potentially drawn to the study due to being interested in physical fitness, or partaking in more physical activity than average levels. Therefore, this sample may not be fully representative of the physical fitness levels of the older population: future studies should consider testing older adults in the community to account for this issue.

To conclude, our study found modest rates of agreement between the PFQ and its individual components, and performance-based measures of balance and strength. While these results are encouraging as significant correlations were observed, the magnitude of these was relatively small and varied based on the performance-based measure under investigation. These findings call into question the interchangeable use of self-report and objective measures in epidemiological research. Nonetheless, the modest rates of convergent validity that we found in our study may be used to encourage the comprehensive assessment of physical fitness through a combination of direct and self-repoted measures.

Chapter 5: Main findings and recommendations for future research

Main findings

The overarching aim of the current work was the exploration of the predominant measures of physical function in ageing research, and the investigation of their role as predictors for cognitive decline and dementia risk. In particular, chapters 2 and 3 investigated whether performance-based measures of physical function were significant predictors of cognitive decline and dementia risk respectively, through the systematic review and meta-analysis of longitudinal cohort studies looking at prospective associations between the variables of interest. As much of the literature uses self-report measures of physical fitness, chapter 4 looked at the validity of self-reported measures in relation to their agreement with performance-based measures of the same construct, and in relation to the share of overlapping variance in a cognitive variable explained by self-reported and objective measures.
Findings reported in chapters 2 and 3 indicated that performance-based physical function measures held strong predictive power for cognitive decline and dementia risk respectively. Several issues surfaced when reviewing evidence from prospective cohort studies, both in terms of the methodology adopted, and of inconsistencies in the reporting of results. These inconsistencies will be further discussed here, along with some recommendations that should be followed by future longitudinal ageing research in order to promote the generation of cohesive findings. Further, the issue of reverse causality in prospective cohort studies will be discussed, with considerations on how this aspect may affect our interpretation of the results, and how it can be partially mitigated for by future research. Finally, this chapter will discuss how these findings, considered within the context of previous research looking at this association, may be used to inform clinical practice and public health policies.
Chapter 4 indicated that performance-based and self-reported measures of physical functioning as assessed through the secondary analysis of data from 3 cross-sectional studies showed modest rates of agreement. Although significant correlations emerged throughout, the related coefficients were only of modest magnitude, which was highlighted as a cause for concern in light of the fact that self-reported and performance-based measures of physical function are used interchangeably in ageing research. The implications of suboptimal convergent validity between self-reported and objective measures will be discussed below, along with recommendations to address this issue in future research. Finally, findings from chapter 4 showed that in a model predicting cognitive scores from age, the PFQ and grip strength, grip strength did not explain any additional variance in the cognitive outcome, that was not already accounted for by the PFQ. We propose that this may be due to grip

strength and PFQ measuring the same construct: thus, when grip strength is entered after the PFQ, the variance that grip strength explains in the cognitive outcome overlaps with the variance that was already explained by the PFQ. So, grip strength does not explain any additional variance that was not already accounted for by the PFQ. These results suggest that the PFQ and handgrip strength are assessing the same construct, and are highly relevant for the validity of the self-reported measure that was used.

Methodological and reporting inconsistencies in prospective cohort studies: promoting common guidelines in ageing research

As discussed in chapters 2 and 3, systematic reviews of the literature looking at prospective associations between physical function and cognitive decline/dementia risk highlighted a number of methodological challenges. These were identified as heterogeneity in the outcome measures used, diverse resolution of different testing measures employed to investigate the same construct, and lack of overall standardised methodological guidelines. Further to this, the statistical techniques used to analyse the data were also highly heterogeneous, as were the reported statistical coefficients of association resulting from such analyses.

There was great diversity in the testing procedures used to establish cognitive outcomes. The majority of the studies retrieved determined cognitive decline as a decline of three points in the MMSE; however, a variety of composite measures derived from different combinations of cognitive tests were also used to measure the same construct. While highlighting an evident heterogeneity of outcome measures, these observations also relate to the sensitivity of the testing procedures: indeed, studies relying on composite measures derived from different tests to determine cognitive outcomes may be more sensitive to detecting cognitive changes than those using the MMSE exclusively. Issues with the resolution (or sensitivity) of the testing procedures used also emerged when considering physical outcomes. This was exemplified by the use of innovative physical assessment methods, which captured aspects of physical variables such as gait stride variability, that were predictive of cognitive impairment, while traditional measures such as simple walking speed were not found to significantly predict cognitive decline. Finally, a lack of standardised guidelines for methodology was evident throughout, from the different length of follow-up periods, the use of different statistical analysis methods resulting in the inconsistent reporting of results, and the use of different combinations of covariates in the analyses. All of these factors made up for significant heterogeneity in the data, which made cross-study comparisons problematic. In light of the wide range of factors that determine such heterogeneity, the adherence to homogeneous, standardised protocols emerged as a pressing issue in ageing research, in order to facilitate cross-study comparisons and promote comprehensive meta-analyses. Based on the abovementioned issues, which were discussed more extensively in the relevant chapters, and on general observations related to the nature of longitudinal research, we will now outline basic guidelines for the design and implementation of longitudinal ageing studies, with a specific focus on the association under investigation.

Designing longitudinal studies

As outlined by Newman's (2010) guidelines for longitudinal ageing studies, there are specific factors that need to be addressed when designing and implementing a longitudinal study. In relation to our association of interest, these factors can be synthesised as (1) the target population, (2) the adequacy of the outcome measures to capture the construct of interest, as well as subtle changes in such construct (3) the statistical analysis that will best encompass the relation between the variables, (4) any potential confounders to the analyses, and (5) the follow-up time required for differences to emerge. Recommended guidelines to address these factors in future longitudinal research will be outlined in the paragraphs below.

Longitudinal studies for the older population

In relation to our target population, the research design should consider the physical and cognitive limitations of older adults, and the study procedure should be developed in a way that mitigates participant burden and reduces rates of attrition. There is evidence showing that both physical and cognitive decline start to be evident after the age of 50 years old (Salthouse, 2009). Therefore, investigating the association between these variables in participants aged 50 or older seems like a reasonable approach. This strategy was adopted by most of the studies that we retrieved; however, when wide age ranges are considered, the stratification of the analysis by age groups is recommended, in order to isolate any age effects. Performance-based physical function tests should be selected in relation to the age of participants, whereby the tasks that participants undergo are not excessively burdensome or demanding. For this purpose, pilot studies where a reduced number of participants attempts the physical tasks are recommended, to identify any floor or ceiling effects and review the procedure accordingly. Finally, in the case of our outcomes of interest, in-person assessments are necessary for the evaluation of performance-based physical function, the diagnosis of dementia, and the determination of cognitive decline. This should be considered during the process of recruitment, because extensive in-person assessments may lead to higher rates of participant attrition, especially in our target population, where higher mortality rates contribute further towards attrition. Therefore, recruitment processes should account for significant attrition over time and, where financially viable, participants should be given the option to be assessed in the community. Further to facilitating participation, this option ensures that the data collected is more closely representative of the general population, as participants with mobility impairments may not always be able to reach testing locations.

Operationalisation of physical and cognitive outcomes

As far as the appropriateness of the outcome measures used, it is essential to operationalise variables in a way that can accurately measure the construct of interest, and that is able to capture the "natural oscillations" of the variables under investigation (Kehr & Kowatsch, 2015). That is to say, the outcome measures used should be able to capture changes that naturally occur in a variable based on a limited number of observations, and should be selected based on the form and speed of change of the variable that is being studied. For this purpose, cognitive measures that are sensitive to detect subtle cognitive changes over relatively short periods of time should be prioritised. With evidence showing that small changes in the MMSE are frequently due to measurement error or practice effects (Hensel et al., 2007), the use of tests such as Category Fluency, Word list- Immediate and Delayed recall, and Trail Making test is recommended instead, as they have shown sensitivity to small cognitive changes (Jutten et al., 2021). Other tests such as the Memory Capacity test and the Short-term Memory Binding test are recommended options, as they show ability to detect subtle cognitive changes, as well as being associated with biomarkers of preclinical Alzheimer's disease (Rentz et al., 2013). Nonetheless, an extensive testing battery is preferable, when this is financially sustainable. For instance, Abe and colleagues (2017) used the "5 Cog" testing battery, which evaluates a range of cognitive functions such as attention, memory, visuospatial function, verbal fluency and reasoning through 5 cognitive tests. The homogeneous adherence to a similar testing protocol may lead to improved testing resolution and greater possibility to make cross-study comparisons in longitudinal research.

In relation to the operationalisation of physical function, the current work has built the case for the use of performance-based measures as opposed to self-report questionnaires. The use of performance-based measures requires in-person assessments, but it should always be encouraged, as it offers an objective picture of physical functioning. As discussed above, community-based assessments of physical performance should always be offered, where possible. Tests such as timed gait, balance tests, handgrip strength, chair-stands test, and timed up and go tests are widely recognised as valid indicators of performance. A testing protocol composed by these tests is easily administrable, and recommended as the gold standard when measuring physical functioning. As highlighted above, chapter 2 showed that recently developed device-based measures of physical performance allowed to observe that specific variables such as gait stride variability showed greater predictive power for cognitive decline than traditional gait speed measures. Similarly, peak VO_2 and fine motor dexterity showed stronger predictive power than traditional cardiorespiratory fitness and strength measures respectively. While these methods are less commonly used, they provide a higher level of accuracy when predicting cognitive outcomes, and are strongly encouraged when the option to incorporate them in study protocols is feasible.

Analysis methods that capture the physical function-cognition association effectively

When deciding on the statistical analysis that bests captures patterns of association, one must always consider the research question under investigation. In the case of our research question, we were looking at patterns of association between baseline physical function and subsequent development of cognitive decline/dementia risk. Linear or logistic regression were the most commonly used analysis techniques, either when the results were considered as a continuous outcome (for instance, continuous rates of cognitive decline associated with poorer physical performance) or as a dichotomous outcome (i.e. presence or absence of cognitive decline/dementia diagnosis) respectively. Based on whether studies considered their results as continuous or discrete outcomes, these results would be reported as either continuous β coefficients, or as odds ratios. The inconsistent methods of reporting that were used in the studies retrieved from our search made it complex to adopt a meta-analytical approach to explain data patterns.

On the other hand, studies looking at the association between changing slopes for two variables (i.e. where both variables were assessed at two or more time points) are becoming increasingly popular, although still not predominant in ageing research, and they use different analysis methods. These studies are becoming predominant as epidemiological research has started to take a life-course perspective, and they usually rely on cross-lagged panel models for their analyses. Cross-lagged panel models are equations used to analyse data with two or more variables repeatedly assessed at two or more time points, in order to investigate the directional effects of one variable on the other at different time points. Linear and logistic regression, and crossed lagged panel models are recommended as the gold standard analyses to investigate this type of association, depending on the research question at hand. However, future research investigating this topic should seek greater consistency in the statistical techniques and the reporting units used, in order to facilitate a cohesive meta-analytical approach. For instance, studies looking at the association between physical function and cognitive decline should report both continuous and discrete outcomes, in order to enable cross-study comparisons. The adherence to common guidelines for reporting is therefore encouraged, although there may be constraints on the analysis type based on the research questions that a study seeks to address.

Establishing basic criteria for covariate adjustment

Confounders are another important factor to consider in longitudinal cohort studies. The set of confounders that may influence the association between physical function and cognition is quite large. The studies that we retrieved from our search adjusted their analyses for covariates that differed widely across studies, and that can be broadly summarised into 6 categories, which include: (1) demographic factors such as age and gender, (2) social factors such as socioeconomic status and education, (3) lifestyle factors such as smoking, diet, social engagement and alcohol intake (4) comorbidities such as cardiovascular disease, hypertension or diabetes, (5) health status, such as depression, anxiety symptoms, body mass index, or cognitive status and (6) genetic factors such as *APOE4* status. For the most part, the studies that were retrieved reported that the adjustment for covariates did not influence the significance of the association of interest. Nonetheless, some of the studies from our search reported that some of the covariates included in their analysis had variable effects on the significance of the research findings. Some studies found that demographic factors explained (Auyeung et al., 2011) or attenuated (Barnes et al., 2003) the association between baseline physical function and cognitive changes. Other studies found that the association between physical function and dementia risk was attenuated by adjusting for baseline cognitive test performance (Dumurgier et al., 2017; Welmer et al., 2014**.** Viscogliosi and colleagues (2017) found that white matter hyperintensities attenuated the association between physical function and cognitive decline.. Comorbidities and baseline cognitive function were also found to attenuate the association between physical function and dementia risk (Doi et al., 2019) or in the case of comorbidities and chronic illness, to explain the association between physical function and cognitive decline (Gale et al., 2014). Given the wide range of evidence retrieved, different combinations of covariates were included in the analyses that we examined, which meant that the statistical coefficients of association that were pooled together in meta-analyses were generated from statistical analyses that were adjusted for different confounders. It may be unrealistic to expect all studies to adopt the same approach, as there are different costs and burdens associated with collecting data on covariates. However, we do recommend that essential criteria for the inclusion of covariates are outlined in longitudinal ageing research, and that the quality of the analyses should be examined against these criteria. We adopted this approach for our analysis, by conducting a quality assessment process which defined the inclusion of age, gender, education and comorbidities as the minimum criteria that should be met for confounders. The adherence to common guidelines for the inclusion of covariates would also contribute to improving the quality of meta-analyses, as the statistical coefficients considered would be generated from analyses controlling for the same confounders.

Defining follow-up duration

In terms of the follow-up period, there are different factors to consider. Because the outcome of interest is most frequently a binary variable at a single time point (i.e., dementia diagnosis/cognitive decline or lack thereof), as opposed to a continuous variable measured over several time points, the assessment of cognition is only necessary at two time points, while in this case the evaluation of physical function is only necessary at baseline. This may not be the case when studies are looking at associations between changing slopes. It should also be considered that financial resources may play a role in defining the length of follow-up, which may also be shaped by the nature of the research question. Thus, the aim of homogeneous follow-up lengths in longitudinal ageing studies is only auspicable. Indeed, follow-up durations varied significantly in the studies that we retrieved; however, defining minimum guidelines for follow-up periods is still beneficial for future longitudinal studies. A minimum follow-up period of three years is recommended for studies investigating cognitive decline as the primary outcome measure; however, when the outcome is dementia diagnosis in absence of cognitive impairment at baseline, it might be necessary to introduce a minimum follow-up period of 5 years to allow for the emerging cognitive changes to be sufficiently evident so as to yield a dementia diagnosis. To provide a more precise, data-informed indication, we averaged the follow-up lengths for the studies where an effect was found for *all* of the physical function variables investigated, or for the only variable investigated, in studies that included a single predictor. For the studies looking at dementia outcomes, the average follow-up for studies showing significant effects for all physical function variables was 6.2 years, whereas for cognitive decline the average length was 4.9 years. Although these averages come from a limited set of studies showing significant effects, they provide a useful indication for future research.

Open data sharing

Further to the adherence to homogeneous protocols and to the recommendations outlined above, it is recommended that in the future, researchers make their data open and freely available. One of the issues that we encountered in our meta-analysis was heterogeneity in the analyses used and the consequent variety of statistical indexes reported, which were often not comparable. Open data sharing, where researchers share their resources by making their data freely and openly available on publishing platforms, would counter this issue by facilitating collaborative comparison. In the case of meta-analyses of longitudinal ageing studies, open data sharing may contribute to mitigating the issue of heterogeneous reporting. Indeed, open data sharing would allow researchers around the world to access datasets and potentially re-analyse existing data with alternative methods, in order to compute comparable statistical coefficients, with the result of easier cross-study comparisons and a deeper investigation of the available data. In a field of research where establishing longitudinal associations could have a great societal impact in terms of dementia prevention and reducing healthcare costs, open data sharing is strongly encouraged moving forward.

Mitigating the issue of reverse causality in longitudinal research

Another issue that emerged from the studies that we retrieved is that of reverse causality. Reverse causality defines an intrinsic problem to observational studies, whereby an association is observed between two variables, but there is a degree of uncertainty as to which variable is responsible for changes in the other. Usually, the research question motivating the study identifies an independent variable, or predictor, and a dependent variable, or outcome, which implies a cause-effect relationship, whereby changes in the predictor determine changes in the outcome. In the case of the relation that we are investigating, different levels of the predictor physical function are thought to determine changes in cognition, or a dementia diagnosis at follow-up. However, longitudinal studies are observational, meaning that there is no manipulation of the independent variable physical function. Instead, different levels of physical function are identified, and cognitive outcomes are studied in association to differential levels of physical performance at baseline. This means that although an association is usually observed, it may be driven by different underlying mechanisms. In the case of our research question, if an association is observed between physical function at baseline and cognitive decline at follow-up, this could be driven by different mechanisms, as proposed by Andrade (2020). It may be that people who develop dementia or experience stronger cognitive decline may also be predisposed to have greater physical inactivity and resulting poorer physical function at baseline, caused by decreased overall activity levels in the early stages of cognitive impairment (Andrade, 2020). It could also be the case that a third factor, such as poor cardiovascular health, which is a predisposing factor for both poor cognitive and physical outcomes, may be driving the association between physical function and cognition (Andrade, 2020). The latter idea also relates to the common cause hypothesis, whereby a common ageing process would account for changes across physical and cognitive functions, resulting in similar trajectories for physical and cognitive performance (Okely et al., 2020).

The issue of reverse causality makes it hard to disentangle the directional relation between physical function and cognition. Unfortunately, reverse causality is an unavoidable caveat of observational studies, related to the nature of this design. Nonetheless, epidemiological research relies strongly on observational studies, as they are highly informative on cognitive trajectories, and allow researchers to investigate patterns of association in exceptionally large samples and over long periods of time, where randomised control trials would be difficult to carry out. Further, although randomised control trials would allow to infer causal relationships, they would also necessitate the enactment of a lengthy intervention aimed at improving physical fitness, which could raise ethical concerns for a potentially inactive control group.

As discussed above, the issue of reverse causality is intrinsic to observational study designs, therefore it cannot be completely addressed. As a result of this, researchers must be cautious not to determine definitive cause-effect relationships when interpreting the results from observational studies. This applies both to the interpretation of findings in the direction of their hypothesis, but also in the direction of reverse causation. For instance, Kivimaki and colleagues (2019) erroneously interpreted the results from their meta-analysis from 19 prospective studies, which showed that physical inactivity was related with greater risk of new onset dementia only during the 10 years before the diagnosis, but not earlier than 10 years prior to dementia onset. Kivimaki et al. (2019) interpreted this association as being attributable to reverse causality, whereby participants may have experienced poor physical functioning leading to physical inactivity only in the 10 years immediately before their dementia diagnosis, as part of the overall slowing that characterises the prodromal stage of dementia, rather than physical inactivity being a predisposing factor for greater dementia risk. Therefore, the authors concluded that physical inactivity was due to the prodromal stages of dementia, and that being physical inactive did not represent a risk factor for dementia. This is a controversial message to convey, as pointed out by Andrade (2020), because it might lead to the incorrect belief that increasing activity levels would not have had a beneficial impact on cognitive trajectories. On the other hand, Andrade (2020) proposes that the observed association could be due to a range of causes other than prodromal dementia, such as the onset of cardiovascular risk factors facilitated by prior physical inactivity during these 10 years, or other lifestyle changes which could have impacted the onset of dementia. Conversely, an association should not be interpreted by researchers as being due to the cause-effect relation that they formulated in their hypothesis, but it ought to be discussed in light of the correlational nature of the relationship between the variables that are observed. In other words, observing an association between physical inactivity and increased dementia risk is not sufficient to determine that poor physical fitness *causes* dementia; instead, studies should direct their focus towards interpreting physical inactivity as a predisposing factor for worse cognitive outcomes, while adopting research strategies to address reverse causality in observational study designs.

While the issue of reverse causality is not completely avoidable in observational studies, there are some considerations that can be made when designing and implementing longitudinal studies, in order to mitigate for it. That is to say, by considering additional aspects of the analysis, the probability that an observable association may be due to reverse causality is less likely. Although not fully resolutive, these strategies contribute towards lending credibility to the conclusions that are drawn from observational studies. One such strategy is becoming increasingly popular in this field of research, and it consists in the utilisation of crossed-lag panel models, which investigate the bi-directional relation between cognitive and physical changes in older adults (Okely et al., 2020). This approach requires the assessment of physical function and cognition at a minimum of two time points, in order to determine whether the slopes of these variables are correlated, and to investigate whether baseline levels of one function are correlated to changes in the other function. By conducting this type of analysis, researchers may make inferences on the direction of the relation between physical function and cognition; however, the directionality of the effect is still not sufficient to determine causality. Nonetheless, if baseline physical function is shown to predict cognitive outcomes, but the opposite is not true, then we can be more confident in making hypotheses on the direction of this association, and substantiate the clinical implications of this finding.

Another strategy to minimise the likelihood of reverse causation impacting the validity of results is the adjustment for cognitive status, and for genetic predisposition to dementia (APOE 4 alleles) at baseline. Although this approach does not rule out reverse causation, the adjustment for these variables, and in particular for the genetic predisposition to dementia, reduces the chances that poor physical function may be related to preclinical dementia-related pathology. Another variable that might be worth adjusting for are neuroimaging correlates of dementia such as white matter hyperintensity and hippocampal volume (Duchowny et al., 2022). Using neuroimaging correlates of dementia as a covariate would contribute towards excluding the possibility that physical function may be caused by preclinical dementia. An additional strategy is the evaluation of neural correlates of dementia in midlife, as well as their association to physical function. This approach helps to exclude the reverse causation hypothesis that incipient dementia may influence physical function, because overt clinical manifestations of dementia are very uncommon before the age of 65 (Duchowny et al., 2022). Although the evaluation of genetic predisposing factors and neural correlates of dementia are probably the most valid tool to reduce reverse causality, there are considerable costs associated with obtaining genetic and neuroimaging data, which should be weighed against the concrete benefit that such data would generate.

When studies follow up participants for a short duration of time, the concern that underlying prodromal dementia may negatively impact physical fitness is even more pressing. To account for this, the exclusion of participants with cognitive impairment at baseline or diagnosed dementia at baseline is also helpful, as it reduces the chances that poor physical function may be attributable to the symptomatology of early stages of dementia. Some studies may also remove participants who experience a dementia diagnosis shortly after baseline, where poor physical function is more likely to have been caused by undiagnosed dementia at baseline. Finally, the issue of reverse causality should be considered within the context of the potential mechanisms that mediate the association between physical fitness and cognitive decline. For instance, the release of age-related inflammatory mediators causes a decrease in muscle mass and strength. Skeletal muscle is known to play a role in the secretion of brain-derived neurotropic factor. Because the age-related release of inflammatory mediators precedes the loss of muscle mass, which in turn leads to a decreased production of brain neurotransmitters leading to impaired cognition, a clear direction can be identified from reduced physical capability to cognitive changes. The fact that these mechanisms are uni-directional and widely documented lends further support to the direction of this association going from physical capability to cognitive changes.

Clinical implications of the association between baseline physical function and later cognitive decline/dementia incidence

Although the strategies discussed above contribute towards reducing the hypothesis of reverse causation, a cause-effect relation between physical function and cognitive outcomes cannot be inferred from observational studies. Thus, the importance of following the recommendations that were previously outlined for the interpretation of findings from observational studies, where variables are not experimentally manipulated, is once again stressed. Nonetheless, the fact that physical inactivity is most frequently observed prior to cognitive impairment still holds clinical significance, even without certainty on cause-effect relationships. This is because in light of this temporal association, poor physical function can be used as a clinical tool for the early identification of subsequent cognitive outcomes, regardless of the cause-effect directionality of the association between these variables.

As a matter of fact, the literature review reported in the current work expands on an already solid body of literature documenting this association (Blondell et al., 2014; Clouston et al., 2013). The outcomes from our reviews, when considered within the broader scope of the literature supporting this association, have important implications for healthcare, policy and clinical practice, as they illustrate the importance of introducing comprehensive assessments of physical function for older adults in clinical practice. The results of physical function evaluations should be used to screen this population for high risk or early signs of cognitive impairment, building on the knowledge that poor performance in the physical function field may be correlated with poor cognitive performance in later life, with the potential to degenerate into a dementia diagnosis. In particular, a protocol assessing different aspects of physical function is recommended, as it should provide an accurate indication of performance. A comprehensive assessment composed of walking speed, balance, cardiovascular performance, strength, and dynamic balance/general mobility has particular relevance for the older population and can be easily implemented in most medical facilities. The gait speed test is suggested to assess walking speed: although this test has different variations, a well-established protocol exists for the 4m distance from a standing start (Studenski et al., 2011). Balance can be initially assessed using low-demanding tasks such as semi-tandem and tandem tests, in order to gauge the individual's skill level. If the participant is able to complete these tasks, the maximum stand time (in seconds) during the one-leg standing test is proposed, which requires a higher level of skill. Cardiovascular performance should be assessed through peak oxygen uptake (VO_2), as it provides a direct indication of oxygen consumption during performance. As far as strength goes, handgrip strength as assessed through a hand-held dynamometer is recommended, as it provides a reliable indication of muscle strength, with well-documented associations with cognitive outcomes (Cui et al., 2021). The timed up and go test (TUG) is recommended to assess dynamic balance and general mobility. The TUG should be implemented due to its documented links with fall risk (Podsialdo et al., 1991) and cognition (Katsumata et al., 2011). Finally, the assessment of an individual's physical functioning status would also benefit from implementing basic activities of daily living (ADLs), defined as the skills needed to carry out one's basic needs such as feeding, toileting and moving (O'Neill & Forman, 2020). These activities are commonly assessed by self-report, or proxy report, but we propose the use of performance-based measures such as the Barthel ADL Index, or the PPT as used by Wilkins et al. (2013) and discussed in chapter 3, which involves daily life activity tasks such as simulated eating or dressing.

This testing protocol would allow to capture the individual's physical function status accurately, and has solid theoretical foundations given the evidence reviewed hereby, aside from additional literature (Clouston et al., 2013; Blondell et al., 2014) documenting the predictive role of these tests for cognitive outcomes. We hereby recommend the implementation of a similar assessment, which should be introduced in clinical practice as part of routine check-ups for adults aged 50 or older. This testing protocol could be easily implemented in clinical practice, due to its practicality, reduced costs and low burden for the individual assessed; at the same time, it holds great potential as an early diagnostic tool. When poor levels of performance are detected, this assessment should be followed up by further cognitive testing. If there is overt evidence of cognitive impairment, individuals should be referred to comprehensive neuropsychological testing or further neuroimaging testing. Conversely, if no evident cognitive decline is detected, individuals should be flagged as at high risk of cognitive impairment, and two-yearly follow up cognitive assessments should be encouraged. Furthermore, older adults with low physical performance levels should be directed to physical fitness interventions, aimed at increasing activity levels and ultimately at improving physical function, with potentially beneficial effects on cognition. Although this strategy may carry additional financial and time costs, it could eventually result in preserving resources, through the early identification of high-risk cognitive trajectories and the prevention of later functional dependence, with the ultimate effect of reducing long-term burden on healthcare systems.

In spite of great research efforts and investments in drug development, there is still no effective treatment addressing the neural substrates of dementia, which makes the identification of alternative methods to detect, delay or prevent cognitive impairment highly relevant. Because we are considering a modifiable lifestyle factor such as physical fitness, the association between physical function and cognition has multiple implications: aside from its potential as an early detection tool for cognitive impairment, this association can be exploited to change cognitive trajectories through interventive approaches. As discussed in chapters 1 and 3, a physically active lifestyle has metabolic and cardiovascular benefits which have been shown to protect the brain from the damaging mechanisms of ageing. The effects of a physically active lifestyle derive from exercise promoting blood flow in the brain, preserving hippocampal volumes, and facilitating the release of neurotransmitters which enable synaptogenesis and neurogenesis. The facilitation of these brain processes, considered in light of our findings, could therefore be used to develop and promote interventions aimed at increasing physical fitness levels starting from midlife, in order to modify cognitive trajectories by enabling these mechanisms. Not only do interventive approaches have the potential to influence cognitive trajectories, but they are also highly beneficial for the improvement of functional capacity. With many older adults reporting high levels of sedentary behaviour, and inability to carry out daily life activities such as climbing stairs or toileting (O'Neill & Forman, 2020), the implementation of preventive approaches to avoid or delay functional decline is of major relevance.

Potential interventions for older adults range from strength training, to cardiovascular endurance or balance-focused programmes. Strength training has been shown to counteract the effects of age-related loss of muscle mass, which has documented links with cognitive impairment through the mediating effect of the inflammatory markers involved in muscle loss (O'Neill & Forman, 2020). As far as interventions focusing on cardiovascular endurance, this type of training has been shown to have beneficial effects on cognition in older adults due to the increased cerebral blood flow that may be facilitated by cardiovascular training, which may in turn mitigate vascular risk factors for dementia such as cardiovascular brain damage or small vessel disease (Ahlskog et al., 2011). Similarly, interventions focusing on balance training, have been shown to have beneficial effects on cognition that are thought to be mediated by the stimulation of the vestibular system, which integrates proprioceptive and visual signals to facilitate balance, although the mechanisms by which balance benefits cognition are not yet fully understood (Rogge et al., 2017).

Any of the intervention types described above has well-documented beneficial effects on cognition, and a positive impact on overall health outcomes. Evidence shows that structured training programmes of as little as 12 weeks, with 3 sessions each week, can already show excellent results on cognitive abilities (Macaulay et al., 2021). Considering the financial impact that the growing rates of cognitive impairment in the elderly have on our society, and the pressure that they place on healthcare systems, the introduction of strategies with the potential to prevent the onset or delay the progression of cognitive decline and dementia has significant societal implications. Interventions offer practical approaches which could contribute towards reducing the significant health care expenditure dedicated to dementia, that is expected to exponentially grow as the population ages. Therefore, we recommend that the knowledge of a beneficial impact of such interventions on cognition is used to promote their implementation in clinical practice, whereby older adults who are screened for physical function as outlined above, and are identified as poor-performing, are referred to targeted intervention programmes by physicians. This approach should ultimately lead older adults to adopt a more physically active lifestyle, and to the improvement of physical function and functional capacity, with the potential to benefit cognitive trajectories on the long term.

The issue of suboptimal convergent validity between self-reported and performance-based measures of physical function

The secondary point that the current work aimed to address is the consistency between self-reported and performance-based measures of physical function, given the widespread use of the former in ageing research. Chapter 4 consisted of the secondary analysis of cross-sectional data from 3 studies using both self-reported and performance-based measures of physical function to investigate their cross-sectional associations with cognition. The results from chapter 4 showed low to moderate rates of agreement between objective and subjective measurements. The correlation coefficients emerging from the analysis were of modest strength, which was identified as a problematic factor, in light of the fact that subjective and objective measures of physical function are frequently used interchangeably in ageing research. At times, self-reported and performance-based measures of physical function may also be pooled together in meta-analyses as compatible predictors, highlighting the problematic nature of this interchangeable use.

The concept of convergent validity was introduced when discussing findings from chapter 4, as it helps to put our findings into a clearer perspective. Convergent validity is commonly discussed when constructing and validating measurement scales, and it refers to the extent to which two different measures capture the same construct (Carlson & Herdman, 2012). During the process of validating a new scale, the correlation of the new scale with already existing measures assessing the same construct is tested. This association is usually regarded as appropriate at the .50 threshold, whereby adequate, although modest levels of convergent validity are met (Abma et al., 2016). Our secondary data analysis found only modest rates of agreement between self-reported and objective measures of physical function, with correlation coefficients ranging from -.66 to -.17, and some of them only marginally approaching the acceptability threshold of .50. Although the values that we found may appear to indicate a certain degree of convergence between subjective and objective physical fitness measures, with relatively small divergence between these two measurement types, there is evidence showing that even a small degree of suboptimal convergent validity may impact the validity of the research results generated from the measure used, and lead to a range of misinterpretations of relative findings (Carlson & Herdman, 2012). In theory, greater rates of agreement between two measures assessing the same construct allow to make confident inferences from research data in support of the research hypotheses. On the other hand, weak rates of agreement between two measurement types of the same construct make the interpretation of research results ambiguous, as they introduce uncertainty as to whether the measures in question appropriately capture the properties of the construct under investigation. As a matter of fact, Carlson and Herdman (2012) propose that the stronger the magnitude of the correlation between two measures is, the greater the likelihood that they will generate comparable research results, and that they are actually measuring the construct accurately. Conversely, the lower the convergence between the two measures, the less likely they are to produce equivalent research findings, and to appropriately reflect properties of the construct of interest. Consequently, when two measures that assess the same construct show weak rates of correlation, this introduces uncertainty as to whether one or both variables are not capturing the construct of interest appropriately, making the interpretation of deriving research findings problematic.

The largest study looking at rates of agreement between subjective and objective measures of physical function conducted to date (Prince et al., 2008) examined correlation coefficients from 187 studies investigating this association. Prince and colleagues (2008) found rates of agreement ranging from .71 to .96, with no clear patterns emerging from the data, and self-reported data generating both higher and lower scores than the objective measures of physical function. Our results from chapter 4 found modest rates of agreement; although there was a clear direction of the association of interest, the range of the magnitude of such association started from as low as .17. In light of the possible ambiguity in the interpretation of data generating from measures with suboptimal convergent validity as discussed above, the inconsistency in these coefficients is somewhat concerning. This is especially true when considering the widespread use of self-reported measures in ageing research, as well as their inclusion in meta-analyses where they are treated as comparable predictors to performance-based measures.

Chapter 4 also included various stepwise regression analyses predicting cognitive scores from age, PFQ scores (overall scores and aerobic, strength and balance sub-scores separately) and grip. Results from these analyses showed that when grip was entered as the last predictor after age and PFQ, it did not explain any additional variance in the cognitive outcome that was not already accounted for by the PFQ. These results suggest that the PFQ and grip are assessing the same construct, and strongly support the validity of this self-reported questionnaire. In fact, if grip does not explain any additional variance in MoCA scores when entered after the PFQ, it is then reasonable to hypothesise that grip and PFQ are measuring the same construct, as they explain the same amount of variance in the cognitive outcome. Thus, when entered after PFQ, the variance explained by grip simply overlaps with that explained by the PFQ, because they in fact capture the same construct. This pattern was further confirmed by additional analyses looking at the individual components of the PFQ separately, which showed that grip did not explain any additional variance when entered after PFQ strength and balance subcomponents, whereby these subcomponents shared measurement properties with grip strength, but it did add significant variance when entered after aerobic conditioning PFQ, which does not share any measurement characteristics with grip. The possibility to implement this type of analysis in scale development and validation processes is promising. Indeed, regression models testing the overlapping variance in an outcome variable that is explained by two variables, one being self-reported and the other its objective counterpart, could lend further credibility to the self-reported measure by demonstrating that it is capturing the same construct as the objective measure. Ideally, a perfect overlap in the variance of an outcome explained by two variables would mean that the self-reported and objective measure are assessing the same construct. We recommended that this type of analysis is implemented in scale development, even at the single item level, whereby individual items are compared to the gold-standard objective measure that the item intends to capture in an iterative manner, until perfect overlap in variance explained is reached. This process would substantiate construct validity and ascertain that the self-reported measure of interest is capturing what it intends to measure. This analysis requires the simultaneous use of self-reported and performance-based measures, as an objective marker of the construct of interest is needed for comparative purposes.

While the use of performance-based measures is desirable, as they more accurately capture the individual's physical status and are not subject to recall bias, this is not always possible in practice. The use of self-reported measures is instead predominant due to their practicality, low cost, and easy accessibility and administration. Generally speaking, there are several instances in epidemiological research where it might be necessary to use alternative measures when assessing a construct, for instance when the gold standard measure is not available, when sensitive data are involved, or when there are financial constraints, and the desirable measure requires the investment of additional resources (Carlson & Herdman, 2012). Therefore, it is necessary to judge when two measures have sufficient convergent validity so as to be used interchangeably. In the case of our construct of interest, the theoretical framework seems to point towards inconsistent associations, which do not support the interchangeable use of self-reported and performance-based measures. Nonetheless, the practical matters outlined above cannot always be addressed, so the use of self-reported measures is at times the more sustainable option. In light of the importance of the findings generated from self-report data on physical functioning, which are used to establish potential risk factors leading to adverse health and cognitive outcomes, as well as to inform policy-making, the use of additional strategies to address suboptimal convergent validity between self-reported and performance-based measures is recommended.

In particular, the use of more than one self-reported measure is encouraged. When the use of performance-based measures is not possible due to financial or timing constraints, the use of two self-reported measures ensures a degree of awareness in the interpretation of results, in case that the two self-reported measures show weak convergent validity. Conversely, if the two self-reported measures show high convergent validity, then these two measures are likely to be measuring the same construct, and are more likely to generate results that are consistent with the desirable measure (Carlson & Herdman, 2012). In the case of our association of interest, when performance-based measures of physical function are not available, both questionnaires and diary-like activity logs could be used as self-reported tools of physical activity. High rates of convergent validity between these two self-reported measures would lend credibility to the results, and increase the likelihood that the two measures are capturing the same construct. However, even in the case that two self-reported measures show high convergent validity, the possibility exists that they are both assessing the same *wrong* construct. Therefore, the adoption of this strategy is only recommendable when there is no access to performance-based measures of physical function. When objective measures are available, the conduction of a pilot study using a combination of performance-based and self-reported measures to assess physical function is encouraged. When the resources available do not allow to test large samples with performance-based measures, piloting a study where convergent validity is tested between the chosen self-reported measure of physical activity and the desired performance-based outcome in a small portion of the sample could lend further credibility to the results, and ensures a solid evidence-based framework for their interpretation.

The recommendations that we outlined in this section ensure that epidemiological studies using self-reported measures to assess physical activity have evidence-based foundations, and are supported by convergent validity analyses. Although the employment of performance-based measures may not always be sustainable, the proposed strategies can partially account for this shortcoming, and solidify findings that are used to inform policies with significant societal value.

Future research

The current body of work points towards a consistent association between physical function and cognition, which is also widely documented in the literature. Building on the evidence that we presented, there are different directions for the next steps in the field of ageing research. The first direction that we recommend is building on the knowledge of an association between greater physical fitness levels and reduced risk of cognitive decline in order to develop physical activity interventions aimed at increasing physical fitness levels. In particular, a direction of future research could test the feasibility and adherence to short-term physical activity and its effects on cognitive levels prior vs post intervention in community-dwelling older adults. This approach would establish whether the longitudinal associations that we reported in our work also translate into a beneficial effect of interventions developed building on those associations, i.e. targeting physical fitness to investigate the impact of manipulating this variable on cognition. Potential interventions may target strength through a three-weekly resistance training programme increasing in difficulty over a 12-week span, or it may target cardiovascular endurance through aerobic training of the same duration. Given the longitudinal associations with reduced cognitive decline and dementia risk that we observed for these physical function variables, we would expect such interventions to have a beneficial impact on cognition.

A secondary direction for future work in ageing research would be the development of a multi-center harmonised longitudinal study protocol, looking at the prospective association between physical function and cognitive decline/dementia risk. In a multi-center study,researchers from different countries could collaborate by collecting data according to standardised methodological and reporting guidelines. In this way, data could be easily pooled into meta-analyses as common outcomes and analysis methods could be pre-established according to standardised protocols. This approach would introduce common guidelines to assess physical function, in order to counter the issue of heterogeneous testing procedures. Further, a multicenter longitudinal investigation would also allow to carefully control for country-specific confounders such as location, dietary and cultural factors, allowing researchers to gain further understanding on the factors moderating the association between physical function and cognitive decline.

A third and final study that could be conducted building on the current work would be an investigation looking at the prospective association between performance-based measures of physical function and cognitive decline in further longitudinal datasets that have not been used to explore this association to date. Extensive performance-based physical function data is contained in the English Longitudinal Study of Ageing (ELSA), a longitudinal study investigating different aspects of ageing in adults in England through the regular collection of data waves over 20 years. This dataset can be merged with the Harmonised Cognitive Assessment Protocol (HCAP), a sub-study of ELSA which explores the prevalence of cognitive impairment and dementia in older adults and includes a wide range of cognitive and neuropsychological measures, which are prone to be affected by cognitive ageing. Merging these two datasets would allow to explore the prospective association between physical function and cognitive decline extensively, given the great variety of physical and cognitive measures included in both datasets. As mentioned, these datasets have not been merged to date, so this would be an interesting and informative approach for future research.

Conclusion

The aim of the current work was to expand on the well-established body of research showing associations between physical fitness and cognitive outcomes in older adults. To this end, prospective cohort studies looking at the association between baseline physical function and subsequent cognitive decline or dementia diagnosis were systematically reviewed in two different chapters, according to the primary outcome. Findings from our reviews pointed towards a predictive role of most of the variables that we considered for both cognitive impairment and dementia. The importance of these findings was discussed, highlighting their value for policy making and healthcare systems. Further, we highlighted some methodological issues which emerged from the studies that we reviewed, such as inconsistencies in the reporting of results, and heterogeneity in the testing procedures used, and made recommendations to improve methodological rigour in future longitudinal research.

The secondary aim of this work was to investigate the consistency between performance-based and self-reported measures of physical function, in light of the widespread use of the latter in this area of research, which emerged as an issue from the searches conducted in our review. To this end, we reported the current agreement on the topic in the literature, and conducted a secondary analysis on data from three studies looking at this association. Our analysis looked at strength, cardiovascular and balance sub-scores of the PFQ in correlation to the specific performance-based measures that they related to, and predicted cognitive outcomes through regression models based on self-reported and objective markers of physical fitness. Our findings showed modest rates of correlation between the objective and subjective measures that we looked at, suggesting that there may be some issues with the interpretation of the results when these measures are considered as equivalent: these issues were discussed in light of the concept of convergent validity in scale development and validation. The overlapping variance in the cognitive outcome explained by self-reported and performance-based measures emerging from our regression analyses was suggestive of shared measurement properties of these measures, and indicated a possible direction for the validation of self-reported measures of physical fitness.

In conclusion, the current work expanded on existing evidence on a highly significant association, which may offer practical solutions to implement in order to address the constantly growing rates of cognitive impairment in the elderly population. Our findings should be considered in light of a well-established body of literature looking at this association, in order to inform the development of strategies to contain the growing dementia incidence. Finally, the recommendations that we outlined throughout our work offer strategies to increase the methodological quality of future longitudinal ageing research looking at this association.

References

Abe, T., Soma, Y., Kitano, N., Jindo, T., Sato, A., Tsunoda, K., Tsuji, T., & Okura, T. (2017). Change in hand dexterity and habitual gait speed reflects cognitive decline over time in healthy older adults: a longitudinal study. *Journal of physical therapy science*, *29*(10), 1737–1741. https://doi.org/10.1589/jpts.29.1737

Abma, I. L., Rovers, M., & van der Wees, P. J. (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing hypotheses and interpreting outcomes. *BMC research notes*, *9*, 226. https://doi.org/10.1186/s13104-016-2034-2

Aburto, J. M., Villavicencio, F., Basellini, U., Kjærgaard, S., & Vaupel, J. W. (2020). Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences*, *117*(10), 5250-5259.

Ahlskog, J. E., Geda, Y. E., Graff-Radford, N. R., & Petersen, R. C. (2011). Physical exercise as a preventive or disease-modifying treatment of dementia and brain aging. *Mayo Clinic proceedings*, *86*(9), 876–884. https://doi.org/10.4065/mcp.2011.0252

Ahlskog, J. E., Geda, Y. E., Graff-Radford, N. R., & Petersen, R. C. (2011). Physical exercise as a preventive or disease-modifying treatment of dementia and brain aging. *Mayo Clinic proceedings*, *86*(9), 876–884. https://doi.org/10.4065/mcp.2011.0252

Ainslie, P. N., Cotter, J. D., George, K. P., Lucas, S., Murrell, C., Shave, R., Thomas, K. N., Williams, M. J., & Atkinson, G. (2008). Elevation in cerebral blood flow velocity with aerobic fitness throughout healthy human ageing. *The Journal of physiology*, *586*(16), 4005–4010. https://doi.org/10.1113/jphysiol.2008.158279

Ainsworth B. E. (2009). How do I measure physical activity in my patients? Questionnaires and objective methods. *British journal of sports medicine*, *43*(1), 6–9. https://doi.org/10.1136/bjsm.2008.052449

Ainsworth, B. E., Caspersen, C. J., Matthews, C. E., Mâsse, L. C., Baranowski, T., & Zhu, W. (2012). Recommendations to improve the accuracy of estimates of physical activity derived from self report. *Journal of physical activity & health*, *9 Suppl 1*(0 1), S76–S84. https://doi.org/10.1123/jpah.9.s1.s76

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., O'Brien, W. L., Bassett, D. R., Jr, Schmitz, K. H., Emplaincourt, P. O., Jacobs, D. R., Jr, & Leon, A. S. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and science in sports and exercise*, *32*(9 Suppl), S498–S504. https://doi.org/10.1097/00005768-200009001-00009

Alfaro-Acha, A., Al Snih, S., Raji, M. A., Kuo, Y. F., Markides, K. S., & Ottenbacher, K. J. (2006). Handgrip strength and cognitive decline in older Mexican Americans. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *61*(8), 859–865. https://doi.org/10.1093/gerona/61.8.859

Alfaro-Acha, A., Al Snih, S., Raji, M. A., Markides, K. S., & Ottenbacher, K. J. (2007). Does 8-foot walk time predict cognitive decline in older Mexicans Americans?. *Journal of the American Geriatrics Society*, *55*(2), 245–251. https://doi.org/10.1111/j.1532-5415.2007.01039.x

Alzheimer's Association (2009). 2009 Alzheimer's disease facts and figures. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *5*(3), 234–270. https://doi.org/10.1016/j.jalz.2009.03.001

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

Amarya, S., Singh, K., & Sabharwal, M. (2018). Ageing Process and Physiological Changes. In G. D'Onofrio, A. Greco, & D. Sancarlo (Eds.), Gerontology. IntechOpen. https://doi.org/10.5772/intechopen.76249

Andrade C. (2020). Reverse Causation, Physical Inactivity, and Dementia. *Indian journal of psychological medicine*, *42*(2), 205–206. https://doi.org/10.4103/IJPSYM.IJPSYM_45_20

Auyeung, T. W., Lee, J. S., Kwok, T., & Woo, J. (2011). Physical frailty predicts future cognitive decline - a four-year prospective study in 2737 cognitively normal older adults. *The journal of nutrition, health & aging*, *15*(8), 690–694. https://doi.org/10.1007/s12603-011-0110-9

Barnes, D. E., & Yaffe, K. (2011). The projected effect of risk factor reduction on Alzheimer's disease prevalence. *The Lancet. Neurology*, *10*(9), 819–828. https://doi.org/10.1016/S1474-4422(11)70072-2

Barnes, D. E., Yaffe, K., Satariano, W. A., & Tager, I. B. (2003). A longitudinal study of cardiorespiratory fitness and cognitive function in healthy older adults. *Journal of the American Geriatrics Society*, *51*(4), 459–465. https://doi.org/10.1046/j.1532-5415.2003.51153.x

Beauchet, O., Annweiler, C., Callisaya, M. L., De Cock, A. M., Helbostad, J. L., Kressig, R. W., Srikanth, V., Steinmetz, J. P., Blumen, H. M., Verghese, J., & Allali, G. (2016). Poor Gait Performance and Prediction of Dementia: Results From a Meta-Analysis. *Journal of the American Medical Directors Association*, *17*(6), 482–490. https://doi.org/10.1016/j.jamda.2015.12.092

Belarbi, K., Burnouf, S., Fernandez-Gomez, F. J., Laurent, C., Lestavel, S., Figeac, M., Sultan, A., Troquier, L., Leboucher, A., Caillierez, R., Grosjean, M. E., Demeyer, D., Obriot, H., Brion, I., Barbot, B., Galas, M. C., Staels, B., Humez, S., Sergeant, N., Schraen-Maschke, S., … Blum, D. (2011). Beneficial effects of exercise in a transgenic mouse model of Alzheimer's disease-like Tau pathology. *Neurobiology of disease*, *43*(2), 486–494. https://doi.org/10.1016/j.nbd.2011.04.022

Best, J. R., Liu-Ambrose, T., Boudreau, R. M., Ayonayon, H. N., Satterfield, S., Simonsick, E. M., Studenski, S., Yaffe, K., Newman, A. B., Rosano, C., & Health, Aging and Body Composition Study (2016). An Evaluation of the Longitudinal, Bidirectional Associations Between Gait Speed and Cognition in Older Women and Men. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *71*(12), 1616–1623. https://doi.org/10.1093/gerona/glw066

Blondell, S. J., Hammersley-Mather, R., & Veerman, J. L. (2014). Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. *BMC public health*, *14*, 510. https://doi.org/10.1186/1471-2458-14-510

Bloom, D. E., & Canning, D. (2007). Commentary: The Preston Curve 30 years on: still sparking fires. *International Journal of Epidemiology*, *36*(3), 498-499.

Boyle, P. A., Buchman, A. S., Wilson, R. S., Leurgans, S. E., & Bennett, D. A. (2009). Association of muscle strength with the risk of Alzheimer disease and the rate of cognitive decline in community-dwelling older persons. *Archives of neurology*, *66*(11), 1339–1344. https://doi.org/10.1001/archneurol.2009.240

Brayne, C. (2007). The elephant in the room—healthy brains in later life, epidemiology and public health. *Nature Reviews Neuroscience*, *8*(3), 233-239.

Breijyeh, Z., & Karaman, R. (2020). Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules (Basel, Switzerland)*, *25*(24), 5789. https://doi.org/10.3390/molecules25245789

Brown, G. C. (2015). Living too long: the current focus of medical research on increasing the quantity, rather than the quality, of life is damaging our health and harming the economy. *EMBO reports*, *16*(2), 137-141.

Buchman, A. S., Boyle, P. A., Wilson, R. S., Tang, Y., & Bennett, D. A. (2007). Frailty is associated with incident Alzheimer's disease and cognitive decline in the elderly. *Psychosomatic medicine*, *69*(5), 483–489. https://doi.org/10.1097/psy.0b013e318068de1d

Buchman, A. S., Boyle, P. A., Yu, L., Shah, R. C., Wilson, R. S., & Bennett, D. A. (2012). Total daily physical activity and the risk of AD and cognitive decline in older adults. *Neurology*, *78*(17), 1323–1329. https://doi.org/10.1212/WNL.0b013e3182535d35

Bullain, S. S., Corrada, M. M., Perry, S. M., & Kawas, C. H. (2016). Sound Body Sound Mind? Physical Performance and the Risk of Dementia in the Oldest-Old: The 90+ Study. *Journal of the American Geriatrics Society*, *64*(7), 1408–1415. https://doi.org/10.1111/jgs.14224

Busse, A. L., Gil, G., Santarém, J. M., & Jacob Filho, W. (2009). Physical activity and cognition in the elderly: A review. *Dementia & neuropsychologia*, *3*(3), 204–208. https://doi.org/10.1590/S1980-57642009DN30300005

Calamia, M., De Vito, A., Bernstein, J. P. K., Weitzner, D. S., Carmichael, O. T., & Keller, J. N. (2018). Pedometer-assessed steps per day as a predictor of cognitive performance in older adults. *Neuropsychology, 32*(8), 941–949. https://doi.org/10.1037/neu0000487

Callisaya, M. L., Launay, C. P., Srikanth, V. K., Verghese, J., Allali, G., & Beauchet, O. (2017). Cognitive status, fast walking speed and walking speed reserve-the Gait and Alzheimer Interactions Tracking (GAIT) study. *GeroScience*, *39*(2), 231–239. https://doi.org/10.1007/s11357-017-9973-y

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17-32.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports*, *100*(2), 126–131.

Cesari, M., Penninx, B. W., Pahor, M., Lauretani, F., Corsi, A. M., Rhys Williams, G., Guralnik, J. M., & Ferrucci, L. (2004). Inflammatory markers and physical performance in older persons: the InCHIANTI study. *The journals of gerontology*. *Series A, Biological sciences and medical sciences*, *59*(3), 242–248. https://doi.org/10.1093/gerona/59.3.m242

Chen ST, Siddarth P, Ercoli LM, Merrill DA, Torres-Gil F, Small GW (2014) Modifiable Risk Factors for Alzheimer Disease and Subjective Memory Impairment across Age Groups. *PLoS ONE* 9(6): e98630. https://doi.org/10.1371/journal.pone.0098630

Chen, S. T., Stevinson, C., Tian, T., Chen, L. J., & Ku, P. W. (2020). Accelerometer-measured daily steps and subjective cognitive ability in older adults: A two-year follow-up study. *Experimental gerontology*, *133*, 110874. https://doi.org/10.1016/j.exger.2020.110874

Chou, M. Y., Nishita, Y., Nakagawa, T., Tange, C., Tomida, M., Shimokata, H., Otsuka, R., Chen, L. K., & Arai, H. (2019). Role of gait speed and grip strength in predicting 10-year cognitive decline among community-dwelling older people. *BMC geriatrics*, *19*(1), 186. https://doi.org/10.1186/s12877-019-1199-7

Christensen, H., Mackinnon, A. J., Korten, A., & Jorm, A. F. (2001). The "common cause hypothesis" of cognitive aging: evidence for not only a common factor but also specific associations of age with vision and grip strength in a cross-sectional analysis. *Psychology and aging*, *16*(4), 588–599. https://doi.org/10.1037//0882-7974.16.4.588

Clouston, S. A., Brewster, P., Kuh, D., Richards, M., Cooper, R., Hardy, R., Rubin, M. S., & Hofer, S. M. (2013). The dynamic relationship between physical function and cognition in longitudinal aging cohorts. *Epidemiologic reviews*, *35*(1), 33–50. https://doi.org/10.1093/epirev/mxs004

Colcombe, S. J., Erickson, K. I., Scalf, P. E., Kim, J. S., Prakash, R., McAuley, E., Elavsky, S., Marquez, D. X., Hu, L., & Kramer, A. F. (2006). Aerobic exercise training increases brain volume in aging humans. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *61*(11), 1166–1170. https://doi.org/10.1093/gerona/61.11.1166

Cooper, R., Kuh, D., Cooper, C., Gale, C. R., Lawlor, D. A., Matthews, F., Hardy, R., & FALCon and HALCyon Study Teams (2011). Objective measures of physical capability and subsequent health: a systematic review. *Age and ageing*, *40*(1), 14–23. https://doi.org/10.1093/ageing/afq117

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and science in sports and exercise*, *35*(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Cui, M., Zhang, S., Liu, Y., Gang, X., & Wang, G. (2021). Grip Strength and the Risk of Cognitive Decline and Dementia: A Systematic Review and Meta-Analysis of Longitudinal Cohort Studies. *Frontiers in aging neuroscience*, *13*, 625551. https://doi.org/10.3389/fnagi.2021.625551

Daviglus, M. L., Bell, C. C., Berrettini, W., Bowen, P. E., Connolly, E. S., Jr, Cox, N. J., Dunbar-Jacob, J. M., Granieri, E. C., Hunt, G., McGarry, K., Patel, D., Potosky, A. L., Sanders-Bush, E., Silberberg, D., & Trevisan, M. (2010). NIH state-of-the-science conference statement: Preventing Alzheimer's disease and cognitive decline. *NIH consensus and state-of-the-science statements*, *27*(4), 1–30.

Deckers, K., van Boxtel, M. P., Schiepers, O. J., de Vugt, M., Muñoz Sánchez, J. L., Anstey, K. J., Brayne, C., Dartigues, J. F., Engedal, K., Kivipelto, M., Ritchie, K., Starr, J. M., Yaffe, K., Irving, K., Verhey, F. R., & Köhler, S. (2015). Target risk factors for dementia prevention: a systematic review and Delphi consensus study on the evidence from observational studies. *International journal of geriatric psychiatry*, *30*(3), 234–246. https://doi.org/10.1002/gps.4245

Demnitz, N., Esser, P., Dawes, H., Valkanova, V., Johansen-Berg, H., Ebmeier, K. P., & Sexton, C. (2016). A systematic review and meta-analysis of cross-sectional studies examining the relationship between mobility and cognition in healthy older adults. *Gait & posture*, *50*, 164–174. https://doi.org/10.1016/j.gaitpost.2016.08.028

Demontiero, O., Vidal, C., & Duque, G. (2012). Aging and bone loss: new insights for the clinician. *Therapeutic advances in musculoskeletal disease*, *4*(2), 61–76. https://doi.org/10.1177/1759720X11430858

Deshpande, N., Metter, E. J., Bandinelli, S., Guralnik, J., & Ferrucci, L. (2009). Gait speed under varied challenges and cognitive decline in older persons: a prospective study. *Age and ageing*, *38*(5), 509–514. https://doi.org/10.1093/ageing/afp093

Dishman, R. K., Washburn, R. A., & Schoeller, D. A. (2001). Measurement of physical activity. *Quest*, *53*(3), 295-309.

Doi, T., Tsutsumimoto, K., Nakakubo, S., Kim, M. J., Kurita, S., Hotta, R., & Shimada, H. (2019). Physical Performance Predictors for Incident Dementia Among Japanese Community-Dwelling Older Adults. *Physical therapy*, *99*(9), 1132–1140. https://doi.org/10.1093/ptj/pzz077

Domingos, C., Costa, P., Santos, N. C., & Pêgo, J. M. (2022). Usability, Acceptability, and Satisfaction of a Wearable Activity Tracker in Older Adults: Observational Study in a Real-Life Context in Northern Portugal. *Journal of Medical Internet Research*, *24*(1), e26652.

Duchowny, K. A., Ackley, S. F., Brenowitz, W. D., Wang, J., Zimmerman, S. C., Caunca, M. R., & Glymour, M. M. (2022). Associations Between Handgrip Strength and Dementia Risk, Cognition, and

Neuroimaging Outcomes in the UK Biobank Cohort Study. *JAMA network open*, *5*(6), e2218314. https://doi.org/10.1001/jamanetworkopen.2022.18314

Dumurgier, J., Artaud, F., Touraine, C., Rouaud, O., Tavernier, B., Dufouil, C., Singh-Manoux, A., Tzourio, C., & Elbaz, A. (2017). Gait Speed and Decline in Gait Speed as Predictors of Incident Dementia. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *72*(5), 655–661. https://doi.org/10.1093/gerona/glw110

Eadie, B. D., Redila, V. A., & Christie, B. R. (2005). Voluntary exercise alters the cytoarchitecture of the adult dentate gyrus by increasing cellular proliferation, dendritic complexity, and spine density. *The Journal of comparative neurology*, *486*(1), 39–47. https://doi.org/10.1002/cne.20493

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

Falck, R. S., Davis, J. C., & Liu-Ambrose, T. (2017). What is the association between sedentary behaviour and cognitive function? A systematic review. *British journal of sports medicine*, *51*(10), 800–811. https://doi.org/10.1136/bjsports-2015-095551

Firth, J., Stubbs, B., Vancampfort, D., Schuch, F., Lagopoulos, J., Rosenbaum, S., & Ward, P. B. (2018). Effect of aerobic exercise on hippocampal volume in humans: A systematic review and meta-analysis. *NeuroImage*, *166*, 230–238. https://doi.org/10.1016/j.neuroimage.2017.11.007

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

Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W. J., Burke, G., McBurnie, M. A., & Cardiovascular Health Study Collaborative Research Group (2001). Frailty in older adults: evidence for a phenotype. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *56*(3), M146–M156. https://doi.org/10.1093/gerona/56.3.m146

Gale, C. R., Allerhand, M., Sayer, A. A., Cooper, C., & Deary, I. J. (2014). The dynamic relationship between cognitive function and walking speed: the English Longitudinal Study of Ageing. *Age (Dordrecht, Netherlands)*, *36*(4), 9682. https://doi.org/10.1007/s11357-014-9682-8

Gemmill, E., Bayles, C. M., McTigue, K., Satariano, W., Sharma, R., & Wilson, J. W. (2011). Factors Associated With Adherence to an Accelerometer Protocol in Older Adults, *Journal of Physical Activity and Health*, 8(8), 1152-1159. Retrieved Sep 11, 2022, from https://journals.humankinetics.com/view/journals/jpah/8/8/article-p1152.xml

Gheysen, F., Poppe, L., DeSmet, A., Swinnen, S., Cardon, G., De Bourdeaudhuij, I., Chastin, S., & Fias, W. (2018). Physical activity to improve cognition in older adults: can physical activity programs enriched with cognitive challenges enhance the effects? A systematic review and meta-analysis. *The international journal of behavioral nutrition and physical activity*, *15*(1), 63. https://doi.org/10.1186/s12966-018-0697-x

Glade M. J. (2010). Oxidative stress and cognitive longevity. *Nutrition (Burbank, Los Angeles County, Calif.)*, *26*(6), 595–603. https://doi.org/10.1016/j.nut.2009.09.014

Hackett, R. A., Davies-Kershaw, H., Cadar, D., Orrell, M., & Steptoe, A. (2018). Walking Speed, Cognitive Function, and Dementia Risk in the English Longitudinal Study of Ageing. *Journal of the American Geriatrics Society*, *66*(9), 1670–1675. https://doi.org/10.1111/jgs.15312

Hagströmer, M., Troiano, R. P., Sjöström, M., & Berrigan, D. (2010). Levels and patterns of objectively assessed physical activity--a comparison between Sweden and the United States. *American journal of epidemiology*, *171*(10), 1055–1064. https://doi.org/10.1093/aje/kwq069

Halloway, S., Wilbur, J., Schoeny, M. E., & Barnes, L. L. (2017). The Relation Between Physical Activity and Cognitive Change in Older Latinos. *Biological research for nursing*, *19*(5), 538–548. https://doi.org/10.1177/1099800417715115

Hamer, M., & Chida, Y. (2009). Physical activity and risk of neurodegenerative disease: a systematic review of prospective evidence. *Psychological medicine*, *39*(1), 3–11. https://doi.org/10.1017/S0033291708003681

Hao, L., Xu, X., Dupre, M. E., Guo, A., Zhang, X., Qiu, L., ... & Gu, D. (2020). Adequate access to healthcare and added life expectancy among older adults in China. *BMC geriatrics*, *20*(1), 1-15.

Harada, C. N., Love, M. C. N., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine*, *29*(4), 737-752.

Healey, E.L., Allen, K.D., Bennell, K., Bowden, J.L., Quicke, J.G. and Smith, R. (2020), Self-Report Measures of Physical Activity. Arthritis Care Res, 72: 717-730. https://doi.org/10.1002/acr.24211

Helmerhorst, H. J., Brage, S., Warren, J., Besson, H., & Ekelund, U. (2012). A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *The international journal of behavioral nutrition and physical activity*, *9*, 103. https://doi.org/10.1186/1479-5868-9-103

Hensel, A., Angermeyer, M. C., & Riedel-Heller, S. G. (2007). Measuring cognitive change in older adults: reliable change indices for the Mini-Mental State Examination. *Journal of neurology, neurosurgery, and psychiatry*, *78*(12), 1298–1303. https://doi.org/10.1136/jnnp.2006.109074

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons.

Ho, S. C., Woo, J., Sham, A., Chan, S. G., & Yu, A. L. (2001). A 3-year follow-up study of social, lifestyle and health predictors of cognitive impairment in a Chinese older cohort. *International journal of epidemiology*, *30*(6), 1389–1396. https://doi.org/10.1093/ije/30.6.1389

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological methods*, *14*(2), 150–164. https://doi.org/10.1037/a0015566

Hofer, S. M., & Sliwinski, M. J. (2001). Understanding Ageing. An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology*, *47*(6), 341–352. https://doi.org/10.1159/000052825

Hollamby, A., Davelaar, E. J., & Cadar, D. (2017). Increased Physical Fitness Is Associated with Higher Executive Functioning in People with Dementia. *Frontiers in public health*, *5*, 346. https://doi.org/10.3389/fpubh.2017.00346

Hötting, K., & Röder, B. (2013). Beneficial effects of physical exercise on neuroplasticity and cognition. *Neuroscience and biobehavioral reviews*, *37*(9 Pt B), 2243–2257. https://doi.org/10.1016/j.neubiorev.2013.04.005

Hsu, C. L., Liang, C. K., Liao, M. C., Chou, M. Y., & Lin, Y. T. (2017). Slow gait speed as a predictor of 1-year cognitive decline in a veterans' retirement community in southern Taiwan. *Geriatrics & gerontology international*, *17 Suppl 1*, 14–19. https://doi.org/10.1111/ggi.13034

Innerd, P., Catt, M., Collerton, J., Davies, K., Trenell, M., Kirkwood, T. B., & Jagger, C. (2015). A comparison of subjective and objective measures of physical activity from the Newcastle 85+ study. *Age and Ageing, 44*(4), 691–694. https://doi.org/10.1093/ageing/afv062

Inzitari, M., Newman, A. B., Yaffe, K., Boudreau, R., de Rekeneire, N., Shorr, R., Harris, T. B., & Rosano, C. (2007). Gait speed predicts decline in attention and psychomotor speed in older adults: the health aging and body composition study. *Neuroepidemiology*, *29*(3-4), 156–162. https://doi.org/10.1159/000111577

Jones, C. J., Rikli, R. E., & Beam, W. C. (1999). A 30-s chair-stand test as a measure of lower body strength in community residing older adults. *Research Quarterly for Exercise and Sport*, 70(2), 113-119.

Jutten, R. J., Sikkes, S., Amariglio, R. E., Buckley, R. F., Properzi, M. J., Marshall, G. A., Rentz, D. M., Johnson, K. A., Teunissen, C. E., Van Berckel, B., Van der Flier, W. M., Scheltens, P., Sperling, R. A., Papp, K. V., Alzheimer Disease Neuroimaging Initiative, & National Alzheimer's Coordinating Center, the Harvard Aging Brain Study, and the Alzheimer Dementia Cohort (2021). Identifying Sensitive Measures of Cognitive Decline at Different Clinical Stages of Alzheimer's Disease. *Journal of the International Neuropsychological Society : JINS*, *27*(5), 426–438. https://doi.org/10.1017/S1355617720000934

Katsumata, Y., Todoriki, H., Yasura, S., & Dodge, H. H. (2011). Timed up and go test predicts cognitive decline in healthy adults aged 80 and older in Okinawa: Keys to Optimal Cognitive Aging (KOCOA) Project. *Journal of the American Geriatrics Society*, *59*(11), 2188–2189. https://doi.org/10.1111/j.1532-5415.2011.03645.x

Kehr, F., & Kowatsch, T. (2015). Quantitative longitudinal research: A review of is literature, and a set of methodological guidelines. *Association for Information Systems*.

Keller, K., & Engelhardt, M. (2014). Strength and muscle mass loss with aging process. Age and strength loss. *Muscles, ligaments and tendons journal*, *3*(4), 346–350.

Kikkert, L., Vuillerme, N., van Campen, J. P., Hortobágyi, T., & Lamoth, C. J. (2016). Walking ability to predict future cognitive decline in old adults: A scoping review. *Ageing research reviews*, *27*, 1–14. https://doi.org/10.1016/j.arr.2016.02.001

Kim, G. R., Sun, J., Han, M., Nam, C. M., & Park, S. (2019). Evaluation of the directional relationship between handgrip strength and cognitive function: the Korean Longitudinal Study of Ageing. *Age and ageing*, *48*(3), 426–432. https://doi.org/10.1093/ageing/afz013

Kim, K. H., Park, S. K., Lee, D. R., & Lee, J. (2019). The Relationship between Handgrip Strength and Cognitive Function in Elderly Koreans over 8 Years: A Prospective Population-Based Study Using Korean Longitudinal Study of Ageing. *Korean journal of family medicine*, *40*(1), 9–15. https://doi.org/10.4082/kjfm.17.0074

Kirk-Sanchez, N. J., & McGough, E. L. (2014). Physical exercise and cognitive performance in the elderly: current perspectives. *Clinical interventions in aging*, *9*, 51–62. https://doi.org/10.2147/CIA.S39506

Kivimäki, M., Singh-Manoux, A., Pentti, J., Sabia, S., Nyberg, S. T., Alfredsson, L., Goldberg, M., Knutsson, A., Koskenvuo, M., Koskinen, A., Kouvonen, A., Nordin, M., Oksanen, T., Strandberg, T., Suominen, S. B., Theorell, T., Vahtera, J., Väänänen, A., Virtanen, M., Westerholm, P., … IPD-Work consortium (2019). Physical inactivity, cardiometabolic disease, and risk of dementia: an individual-participant meta-analysis. *BMJ (Clinical research ed.)*, *365*, l1495. https://doi.org/10.1136/bmj.l1495

Kobayashi-Cuya, K. E., Sakurai, R., Sakuma, N., Suzuki, H., Yasunaga, M., Ogawa, S., Takebayashi, T., & Fujiwara, Y. (2018). Hand dexterity, not handgrip strength, is associated with executive function in Japanese community-dwelling older adults: a cross-sectional study. *BMC geriatrics*, *18*(1), 192. https://doi.org/10.1186/s12877-018-0880-6

Kobayashi-Cuya, K.E., Sakurai, R., Sakuma, N. *et al*. (2018). Hand dexterity, not handgrip strength, is associated with executive function in Japanese community-dwelling older adults: a cross-sectional study. *BMC Geriatr*, 192(18).

Kowalski, K., Rhodes, R., Naylor, P. J., Tuokko, H., & MacDonald, S. (2012). Direct and indirect measurement of physical activity in older adults: a systematic review of the literature. *The international journal of behavioral nutrition and physical activity*, *9*, 148. https://doi.org/10.1186/1479-5868-9-148

Ku, P. W., Liu, Y. T., Lo, M. K., Chen, L. J., & Stubbs, B. (2017). Higher levels of objectively measured sedentary behavior is associated with worse cognitive ability: Two-year follow-up study in community-dwelling older adults. *Experimental gerontology*, *99*, 110–114. https://doi.org/10.1016/j.exger.2017.09.014

Kuh, D., Richards, M., Cooper, R., Hardy, R., & Ben-Shlomo, Y. (2014). Life course epidemiology, ageing research and maturing cohort studies: a dynamic combination for understanding healthy ageing. *A life course approach to healthy ageing*, 3-15.

Lee, A. T., Richards, M., Chan, W. C., Chiu, H. F., Lee, R. S., & Lam, L. C. (2015). Poor Balance as a Noncognitive Predictor of Incident Dementia. *Journal of the American Geriatrics Society*, *63*(8), 1701–1702. https://doi.org/10.1111/jgs.13587

Lee, J. E., Shin, D. W., Jeong, S. M., Son, K. Y., Cho, B., Yoon, J. L., Park, B. J., Kwon, I. S., Lee, J., & Kim, S. (2018). Association Between Timed Up and Go Test and Future Dementia Onset. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *73*(9), 1238–1243. https://doi.org/10.1093/gerona/glx261

Leinonen, A. M., Ahola, R., Kulmala, J., Hakonen, H., Vähä-Ypyä, H., Herzig, K. H., Auvinen, J., Keinänen-Kiukaanniemi, S., Sievänen, H., Tammelin, T. H., Korpelainen, R., & Jämsä, T. (2017). Measuring Physical Activity in Free-Living Conditions-Comparison of Three Accelerometry-Based Methods. *Frontiers in physiology*, *7*, 681. https://doi.org/10.3389/fphys.2016.00681

Lewis, F., Karlsberg Schaffer, S., Sussex, J., O'Neill, P., & Cockcroft, L. (2014). The trajectory of dementia in the UK-making a difference. *Office of Health Economics Consulting Reports*, *2013*, 1-55.

Liang, K. Y., Mintun, M. A., Fagan, A. M., Goate, A. M., Bugg, J. M., Holtzman, D. M., Morris, J. C., & Head, D. (2010). Exercise and Alzheimer's disease biomarkers in cognitively normal older adults. *Annals of neurology*, *68*(3), 311–318. https://doi.org/10.1002/ana.22096

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

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., ... & Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, *396*(10248), 413-446.

Lohne-Seiler, H., Hansen, B. H., Kolle, E., & Anderssen, S. A. (2014). Accelerometer-determined physical activity and self-reported health in a population of older adults (65-85 years): a cross-sectional study. *BMC public health*, *14*, 284. https://doi.org/10.1186/1471-2458-14-284

Lyden, K., Kozey, S. L., Staudenmeyer, J. W., & Freedson, P. S. (2011). A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *European journal of applied physiology*, *111*(2), 187–201. https://doi.org/10.1007/s00421-010-1639-8

Macaulay, T. R., Pa, J., Kutch, J. J., Lane, C. J., Duncan, D., Yan, L., & Schroeder, E. T. (2021). 12 weeks of strength training improves fluid cognition in older adults: A nonrandomized pilot trial. *PloS one*, *16*(7), e0255018. https://doi.org/10.1371/journal.pone.0255018

MacDonald, S. W., Hundza, S., Love, J. A., DeCarlo, C. A., Halliday, D. W., Brewster, P. W., Lukyn, T. V., Camicioli, R., & Dixon, R. A. (2017). Concurrent Indicators of Gait Velocity and Variability Are Associated with 25-Year Cognitive Change: A Retrospective Longitudinal Investigation. *Frontiers in aging neuroscience*, *9*, 17. https://doi.org/10.3389/fnagi.2017.00017

Mandolesi, L., Polverino, A., Montuori, S., Foti, F., Ferraioli, G., Sorrentino, P., & Sorrentino, G. (2018). Effects of Physical Exercise on Cognitive Functioning and Wellbeing: Biological and Psychological Benefits. *Frontiers in psychology*, *9*, 509. https://doi.org/10.3389/fpsyg.2018.00509

Marioni, R. E., Valenzuela, M. J., van den Hout, A., Brayne, C., Matthews, F. E., & MRC Cognitive Function and Ageing Study (2012). Active cognitive lifestyle is associated with positive cognitive health transitions and compression of morbidity from age sixty-five. *PloS one*, *7*(12), e50940. https://doi.org/10.1371/journal.pone.0050940

Martinowich, K., & Lu, B. (2008). Interaction between BDNF and serotonin: role in mood disorders. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *33*(1), 73–83. https://doi.org/10.1038/sj.npp.1301571

Massion J. (1998). Postural control systems in developmental perspective. *Neuroscience and biobehavioral reviews*, *22*(4), 465–472. https://doi.org/10.1016/s0149-7634(97)00031-6

McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., & Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science (New York, N.Y.)*, *276*(5318), 1560–1563. https://doi.org/10.1126/science.276.5318.1560

McGough, E. L., Kelly, V. E., Logsdon, R. G., McCurry, S. M., Cochrane, B. B., Engel, J. M., & Teri, L. (2011). Associations between physical performance and executive function in older adults with mild cognitive impairment: gait speed and the timed "up & go" test. *Physical therapy*, *91*(8), 1198–1207. https://doi.org/10.2522/ptj.20100372

Meijer, E. P., Goris, A. H., Wouters, L., & Westerterp, K. R. (2001). Physical inactivity as a determinant of the physical activity level in the elderly. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, *25*(7), 935–939. https://doi.org/10.1038/sj.ijo.0801644

Meunier, C. C., Smit, E., Fitzpatrick, A. L., & Odden, M. C. (2021). Balance and cognitive decline in older adults in the cardiovascular health study. *Age and ageing*, *50*(4), 1342–1348. https://doi.org/10.1093/ageing/afab038

Michaelson D. M. (2014). APOE ε4: the most prevalent yet understudied risk factor for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *10*(6), 861–868. https://doi.org/10.1016/j.jalz.2014.06.015

Mielke, M. M., Roberts, R. O., Savica, R., Cha, R., Drubach, D. I., Christianson, T., Pankratz, V. S., Geda, Y. E., Machulda, M. M., Ivnik, R. J., Knopman, D. S., Boeve, B. F., Rocca, W. A., & Petersen, R. C. (2013). Assessing the temporal relationship between cognition and gait: slow gait predicts cognitive decline in the Mayo Clinic Study of Aging. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *68*(8), 929–937. https://doi.org/10.1093/gerona/gls256

Miranda, M., Morici, J. F., Zanoni, M. B., & Bekinschtein, P. (2019). Brain-Derived Neurotrophic Factor: A Key Molecule for Memory in the Healthy and the Pathological Brain. *Frontiers in cellular neuroscience*, *13*, 363. https://doi.org/10.3389/fncel.2019.00363

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Montero-Odasso, M., Speechley, M., Muir-Hunter, S. W., Sarquis-Adamson, Y., Sposato, L. A., Hachinski, V., Borrie, M., Wells, J., Black, A., Sejdić, E., Bherer, L., Chertkow, H., & Canadian Gait and Cognition Network (2018). Motor and Cognitive Trajectories Before Dementia: Results from Gait and Brain Study. *Journal of the American Geriatrics Society*, *66*(9), 1676–1683. https://doi.org/10.1111/jgs.15341

Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, Currie M, Qureshi R, Mattis P, Lisy K, Mu P-F. (2017). Chapter 7: Systematic reviews of etiology and risk . In: Aromataris E, Munn Z (Editors). *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute. Available from https://reviewersmanual.joannabriggs.org/

Morley, J. E., Baumgartner, R. N., Roubenoff, R., Mayer, J., & Nair, K. S. (2001). Sarcopenia. *The Journal of laboratory and clinical medicine*, *137*(4), 231–243. https://doi.org/10.1067/mlc.2001.113504

Murer, M. G., Yan, Q., & Raisman-Vozari, R. (2001). Brain-derived neurotrophic factor in the control human brain, and in Alzheimer's disease and Parkinson's disease. *Progress in neurobiology*, *63*(1), 71–124. https://doi.org/10.1016/s0301-0082(00)00014-9

Murman D. L. (2015). The Impact of Age on Cognition. *Seminars in hearing*, *36*(3), 111–121. https://doi.org/10.1055/s-0035-1555115

Muscari, A., Spiller, I., Bianchi, G., Fabbri, E., Forti, P., Magalotti, D., Pandolfi, P., Zoli, M., & Pianoro Study Group (2018). Predictors of cognitive impairment assessed by Mini Mental State Examination in community-dwelling older adults: relevance of the step test. *Experimental gerontology*, *108*, 69–76. https://doi.org/10.1016/j.exger.2018.03.021

Naftali Raz, Ulman Lindenberger, Karen M. Rodrigue, Kristen M. Kennedy, Denise Head, Adrienne Williamson, Cheryl Dahle, Denis Gerstorf, James D. Acker, Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers, *Cerebral Cortex*, Volume 15, Issue 11, November 2005, Pages 1676–1689, https://doi.org/10.1093/cercor/bhi044

National Health System (NHS). (2021). *Overview - Alzheimer's Disease*. NHS, accessed from https://www.nhs.uk/conditions/alzheimers-disease/

Newman A. B. (2010). An overview of the design, implementation, and analyses of longitudinal studies on aging. *Journal of the American Geriatrics Society*, *58 Suppl 2*(Suppl 2), S287–S291. https://doi.org/10.1111/j.1532-5415.2010.02916.x

Nucci, M., Mapelli, D., & Mondini, S. (2012). Cognitive Reserve Index questionnaire (CRIq): a new instrument for measuring cognitive reserve. *Aging clinical and experimental research*, *24*(3), 218–226. https://doi.org/10.3275/7800

O'Neill, D., & Forman, D. E. (2020). The importance of physical function as a clinical outcome: Assessment and enhancement. *Clinical cardiology*, *43*(2), 108–117. https://doi.org/10.1002/clc.23311

Ogoh, S., Tsukamoto, H., Hirasawa, A., Hasegawa, H., Hirose, N., & Hashimoto, T. (2014). The effect of changes in cerebral blood flow on cognitive function during exercise. *Physiological reports*, *2*(9), e12163. https://doi.org/10.14814/phy2.12163

Ogonowska-Slodownik, A., Morgulec-Adamowicz, N., Geigle, P.R. *et al*. Objective and Self-reported Assessment of Physical Activity of Women over 60 Years Old. *Ageing Int* **47**, 307–320 (2022). https://doi.org/10.1007/s12126-021-09423-z

Okely, J. A., & Deary, I. J. (2020). Associations Between Declining Physical and Cognitive Functions in the Lothian Birth Cohort 1936. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *75*(7), 1393–1402. https://doi.org/10.1093/gerona/glaa023

Padilla Colón, C. J., Molina-Vicenty, I. L., Frontera-Rodríguez, M., García-Ferré, A., Rivera, B. P., Cintrón-Vélez, G., & Frontera-Rodríguez, S. (2018). Muscle and Bone Mass Loss in the Elderly Population: Advances in diagnosis and treatment. *Journal of Biomedicine (Sydney, NSW)*, *3*, 40–49. https://doi.org/10.7150/jbm.23390

Parker, S. J., Strath, S. J., & Swartz, A. M. (2008). Physical activity measurement in older adults: relationships with mental health. *Journal of aging and physical activity*, *16*(4), 369–380. https://doi.org/10.1123/japa.16.4.369

Paolillo, E. W., Lee, S. Y., VandeBunte, A., Djukic, N., Fonseca, C., Kramer, J. H., & Casaletto, K. B. (2022). Wearable Use in an Observational Study Among Older Adults: Adherence, Feasibility, and Effects of Clinicodemographic Factors. *Frontiers in Digital Health*, *4*.

Pathan, S. S., Gottesman, R. F., Mosley, T. H., Knopman, D. S., Sharrett, A. R., & Alonso, A. (2011). Association of lung function with cognitive decline and dementia: the Atherosclerosis Risk in Communities (ARIC) Study. *European journal of neurology*, *18*(6), 888–898. https://doi.org/10.1111/j.1468-1331.2010.03340.x

Pedersen, B. K., & Febbraio, M. A. (2012). Muscles, exercise and obesity: skeletal muscle as a secretory organ. *Nature reviews. Endocrinology*, *8*(8), 457–465. https://doi.org/10.1038/nrendo.2012.49
Peters R. Ageing and the brain. Postgrad Med J. 2006 Feb;82(964):84-8. doi: 10.1136/pgmj.2005.036665.

Podsiadlo, D., & Richardson, S. (1991). The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, *39*(2), 142–148. https://doi.org/10.1111/j.1532-5415.1991.tb01616.x

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *The international journal of behavioral nutrition and physical activity*, *5*, 56. https://doi.org/10.1186/1479-5868-5-56

Proust-Lima, C., Amieva, H., Dartigues, J. F., & Jacqmin-Gadda, H. (2007). Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. *American journal of epidemiology*, *165*(3), 344–350. https://doi.org/10.1093/aje/kwk017

Quan, M., Xun, P., Chen, C., Wen, J., Wang, Y., Wang, R., Chen, P., & He, K. (2017). Walking Pace and the Risk of Cognitive Decline and Dementia in Elderly Populations: A Meta-analysis of Prospective Cohort Studies. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *72*(2), 266–270. https://doi.org/10.1093/gerona/glw121

Rappuoli, R., Pizza, M., Del Giudice, G., & De Gregorio, E. (2014). Vaccines, new opportunities for a new society. *Proceedings of the National Academy of Sciences*, *111*(34), 12288-12293.

Rechel, B., Doyle, Y., Grundy, E., & McKee, M. (2009). How can health systems respond to population ageing?.

Rentz, D. M., Parra Rodriguez, M. A., Amariglio, R., Stern, Y., Sperling, R., & Ferris, S. (2013). Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. *Alzheimer's research & therapy*, *5*(6), 58. https://doi.org/10.1186/alzrt222

Ritchie, S. J., Tucker-Drob, E. M., Starr, J. M., & Deary, I. J. (2016). Do Cognitive and Physical Functions Age in Concert from Age 70 to 76? Evidence from the Lothian Birth Cohort 1936. *The Spanish journal of psychology*, *19*, E90. https://doi.org/10.1017/sjp.2016.85

Rogge, A. K., Röder, B., Zech, A., Nagel, V., Hollander, K., Braumann, K. M., & Hötting, K. (2017). Balance training improves memory and spatial cognition in healthy adults. *Scientific reports*, *7*(1), 5661. https://doi.org/10.1038/s41598-017-06071-9

Rolandi, E., Zaccaria, D., Vaccaro, R. *et al*. Estimating the potential for dementia prevention through modifiable risk factors elimination in the real-world setting: a population-based study. *Alz Res Therapy* **12**, 94 (2020). https://doi.org/10.1186/s13195-020-00661-y

Russ, T. C., Starr, J. M., Stamatakis, E., Kivimäki, M., & Batty, G. D. (2015). Pulmonary function as a risk factor for dementia death: an individual participant meta-analysis of six UK general population cohort studies. *Journal of Epidemiology and Community Health (1979-)*, *69*(6), 550–556. http://www.jstor.org/stable/44018461

Sachdev, P. S., Wen, W., Christensen, H., & Jorm, A. F. (2005). White matter hyperintensities are related to physical disability and poor motor function. *Journal of neurology, neurosurgery, and psychiatry*, *76*(3), 362–367. https://doi.org/10.1136/jnnp.2004.042945

Sallis, J.F., Haskell, W.L., Wood , P.D., Fortmann, S.P., Rogers, T., Blair, S.N., and Paffenbarger, R. (1985). Physical activity assessment methodology in the Five City Project. American Journal of Epidemiology, 121, 91-106.

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

Sattler, C., Erickson, K. I., Toro, P., & Schröder, J. (2011). Physical fitness as a protective factor for cognitive impairment in a prospective population-based study in Germany. *Journal of Alzheimer's disease : JAD*, *26*(4), 709–718. https://doi.org/10.3233/JAD-2011-110548

Schaller, A., Rudolf, K., Dejonghe, L., Grieben, C., & Froboese, I. (2016). Influencing Factors on the Overestimation of Self-Reported Physical Activity: A Cross-Sectional Analysis of Low Back Pain Patients and Healthy Controls. *BioMed research international*, *2016*, 1497213. https://doi.org/10.1155/2016/1497213

Scholes, S., & Neave, A. (2017). Health Survey for England 2016: Physical activity in adults. *Leeds: Health and Social Care Information Centre*.

Sesso H. D. (2007). Invited commentary: a challenge for physical activity epidemiology. *American journal of epidemiology*, *165*(12), 1351–1355. https://doi.org/10.1093/aje/kwm093

Shephard R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British journal of sports medicine*, *37*(3), 197–206. https://doi.org/10.1136/bjsm.37.3.197

Sibbett, R. A., Russ, T. C., Allerhand, M., Deary, I. J., & Starr, J. M. (2018). Physical fitness and dementia risk in the very old: a study of the Lothian Birth Cohort 1921. *BMC psychiatry*, *18*(1), 285. https://doi.org/10.1186/s12888-018-1851-3

Singh-Manoux, A., & Kivimäki, M. (2010). The importance of cognitive aging for understanding dementia. *Age (Dordrecht, Netherlands)*, *32*(4), 509–512. https://doi.org/10.1007/s11357-010-9147-7

Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., Ferrie, J. E., & Dugravot, A. (2012). Timing of onset of cognitive decline: results from Whitehall II prospective cohort study. *BMJ (Clinical research ed.)*, *344*, d7622. https://doi.org/10.1136/bmj.d7622

Sirard, J. R., & Pate, R. R. (2001). Physical activity assessment in children and adolescents. *Sports medicine (Auckland, N.Z.)*, *31*(6), 439–454. https://doi.org/10.2165/00007256-200131060-00004

Smith, P. F., & Zheng, Y. (2013). From ear to uncertainty: vestibular contributions to cognitive function. *Frontiers in integrative neuroscience*, *7*, 84. https://doi.org/10.3389/fnint.2013.00084

Sofi, F., Abbate, R., Gensini, G. F., & Casini, A. (2010). Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. *The American journal of clinical nutrition*, *92*(5), 1189–1196. https://doi.org/10.3945/ajcn.2010.29673

Stern Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society : JINS*, *8*(3), 448–460.
Stijntjes, M., Aartsen, M. J., Taekema, D. G., Gussekloo, J., Huisman, M., Meskers, C., de Craen, A., & Maier, A. B. (2017). Temporal Relationship Between Cognitive and Physical Performance in Middle-Aged to Oldest Old People. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *72*(5), 662–668. https://doi.org/10.1093/gerona/glw133

Stubbs, B., Chen, L. J., Chang, C. Y., Sun, W. J., & Ku, P. W. (2017). Accelerometer-assessed light physical activity is protective of future cognitive ability: A longitudinal study among community dwelling older adults. *Experimental gerontology*, *91*, 104–109. https://doi.org/10.1016/j.exger.2017.03.003

Studenski, S., Perera, S., Patel, K., Rosano, C., Faulkner, K., Inzitari, M., Brach, J., Chandler, J., Cawthon, P., Connor, E. B., Nevitt, M., Visser, M., Kritchevsky, S., Badinelli, S., Harris, T., Newman, A. B., Cauley, J., Ferrucci, L., & Guralnik, J. (2011). Gait speed and survival in older adults. *JAMA*, *305*(1), 50–58. https://doi.org/10.1001/jama.2010.1923

Suzuki, H., Kawai, H., Hirano, H., Yoshida, H., Ihara, K., Kim, H., Chaves, P. H., Minami, U., Yasunaga, M., Obuchi, S., & Fujiwara, Y. (2015). One-Year Change in the Japanese Version of the Montreal Cognitive Assessment Performance and Related Predictors in Community-Dwelling Older Adults. *Journal of the American Geriatrics Society*, *63*(9), 1874–1879. https://doi.org/10.1111/jgs.13595

Taekema, D. G., Ling, C. H., Kurrle, S. E., Cameron, I. D., Meskers, C. G., Blauw, G. J., Westendorp, R. G., de Craen, A. J., & Maier, A. B. (2012). Temporal relationship between handgrip strength and cognitive performance in oldest old people. *Age and ageing*, *41*(4), 506–512. https://doi.org/10.1093/ageing/afs013

Taniguchi, Y., Kitamura, A., Seino, S., Murayama, H., Amano, H., Nofuji, Y., Nishi, M., Yokoyama, Y., Shinozaki, T., Yokota, I., Matsuyama, Y., Fujiwara, Y., & Shinkai, S. (2017). Gait Performance Trajectories and Incident Disabling Dementia Among Community-Dwelling Older Japanese. *Journal of the American Medical Directors Association*, *18*(2), 192.e13–192.e20. https://doi.org/10.1016/j.jamda.2016.10.015

Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *The Journal of Clinical Psychiatry, 48*(8), 314–318.

Trost, S. G., & O'Neil, M. (2013). Clinical use of objective measures of physical activity. *British journal of sports medicine*, *48*(3), 178–181. https://doi.org/10.1136/bjsports-2013-093173

Tudor-Locke, C., Craig, C. L., Brown, W. J., Clemes, S. A., De Cocker, K., Giles-Corti, B., Hatano, Y., Inoue, S., Matsudo, S. M., Mutrie, N., Oppert, J. M., Rowe, D. A., Schmidt, M. D., Schofield, G. M., Spence, J. C., Teixeira, P. J., Tully, M. A., & Blair, S. N. (2011). How many steps/day are enough? For adults. *The international journal of behavioral nutrition and physical activity*, *8*, 79. https://doi.org/10.1186/1479-5868-8-79

Vallianatou, T., Shariatgorji, M., Nilsson, A., Fridjonsdottir, E., Källback, P., Schintu, N., Svenningsson, P., & Andrén, P. E. (2019). Molecular imaging identifies age-related attenuation of acetylcholine in retrosplenial cortex in response to acetylcholinesterase inhibition. *Neuropsychopharmacology : official* *publication of the American College of Neuropsychopharmacology*, *44*(12), 2091–2098. https://doi.org/10.1038/s41386-019-0397-5

Van Gelder, B. M., Tijhuis, M. A., Kalmijn, S., Giampaoli, S., Nissinen, A., & Kromhout, D. (2004). Physical activity in relation to cognitive decline in elderly men: the FINE Study. *Neurology*, *63*(12), 2316–2321. https://doi.org/10.1212/01.wnl.0000147474.29994.35

van Sluijs, E. M., Griffin, S. J., & van Poppel, M. N. (2007). A cross-sectional study of awareness of physical activity: associations with personal, behavioral and psychosocial factors. *The international journal of behavioral nutrition and physical activity*, *4*, 53. https://doi.org/10.1186/1479-5868-4-53

Verghese, J., Wang, C., Lipton, R. B., Holtzer, R., & Xue, X. (2007). Quantitative gait dysfunction and risk of cognitive decline and dementia. *Journal of neurology, neurosurgery, and psychiatry*, *78*(9), 929–935. https://doi.org/10.1136/jnnp.2006.106914

Veronese, N., Stubbs, B., Trevisan, C., Bolzetta, F., De Rui, M., Solmi, M., Sartori, L., Musacchio, E., Zambon, S., Perissinotto, E., Crepaldi, G., Manzato, E., & Sergi, G. (2016). What physical performance measures predict incident cognitive decline among intact older adults? A 4.4year follow up study. *Experimental gerontology*, *81*, 110–118. https://doi.org/10.1016/j.exger.2016.05.008

Viscogliosi, G., Di Bernardo, M. G., Ettorre, E., & Chiriac, I. M. (2017). Handgrip Strength Predicts Longitudinal Changes in Clock Drawing Test Performance. An Observational Study in a Sample of Older Non-Demented Adults. *The journal of nutrition, health & aging*, *21*(5), 593–596. https://doi.org/10.1007/s12603-016-0816-9

Wadsworth, P. A., Chen, N. W., Raji, M., Markides, K. S., & Downer, B. (2020). Mobility but Not Balance Limitations Are Associated With Cognitive Decline among Older Hispanics. *Gerontology & geriatric medicine*, *6*, 2333721420947952. https://doi.org/10.1177/2333721420947952

Wang, L., Larson, E. B., Bowen, J. D., & van Belle, G. (2006). Performance-based physical function and future dementia in older people. *Archives of internal medicine*, *166*(10), 1115–1120. https://doi.org/10.1001/archinte.166.10.1115

Watkinson, C., van Sluijs, E. M., Sutton, S., Hardeman, W., Corder, K., & Griffin, S. J. (2010). Overestimation of physical activity level is associated with lower BMI: a cross-sectional analysis. *International journal of behavioral nutrition and physical activity*, *7*(1), 1-9.

Welmer, A. K., Rizzuto, D., Qiu, C., Caracciolo, B., & Laukka, E. J. (2014). Walking speed, processing speed, and dementia: a population-based longitudinal study. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *69*(12), 1503–1510. https://doi.org/10.1093/gerona/glu047

Wilkins, C. H., Roe, C. M., Morris, J. C., & Galvin, J. E. (2013). Mild physical impairment predicts future diagnosis of dementia of the Alzheimer's type. *Journal of the American Geriatrics Society*, *61*(7), 1055–1059. https://doi.org/10.1111/jgs.12255

World Health Organization (WHO). (2020). guidelines on physical activity and sedentary behaviour: web annex: evidence profiles."

World Health Organization. (2015). *World health statistics 2015*. World Health Organization.

World Health Organization. (2019). *World health statistics overview 2019: monitoring health for the SDGs, sustainable development goals* (No. WHO/DAD/2019.1). World Health Organization.

Yaffe, K., Fiocco, A. J., Lindquist, K., Vittinghoff, E., Simonsick, E. M., Newman, A. B., Satterfield, S., Rosano, C., Rubin, S. M., Ayonayon, H. N., Harris, T. B., & Health ABC Study (2009). Predictors of maintaining cognitive function in older adults: the Health ABC study. *Neurology*, *72*(23), 2029–2035. https://doi.org/10.1212/WNL.0b013e3181a92c36

Yuki, A., Lee, S., Kim, H., Kozakai, R., Ando, F., & Shimokata, H. (2012). Relationship between physical activity and brain atrophy progression. *Medicine and science in sports and exercise*, *44*(12), 2362–2368. https://doi.org/10.1249/MSS.0b013e3182667d1d

Zaninotto, P., Batty, G. D., Allerhand, M., & Deary, I. J. (2018). Cognitive function trajectories and their determinants in older people: 8 years of follow-up in the English Longitudinal Study of Ageing. *Journal of epidemiology and community health*, *72*(8), 685–694. https://doi.org/10.1136/jech-2017-210116

Zhang, Z., Giordani, B., Margulis, A., & Chen, W. (2022). Efficacy and acceptability of using wearable activity trackers in older adults living in retirement communities: a mixed method study. *BMC geriatrics*, *22*(1), 231.

Zhu, W., Wadley, V. G., Howard, V. J., Hutto, B., Blair, S. N., & Hooker, S. P. (2017). Objectively Measured Physical Activity and Cognitive Function in Older Adults. *Medicine and science in sports and exercise*, *49*(1), 47–53. https://doi.org/10.1249/MSS.0000000000001079

Zimmermann-Sloutskis, D., Wanner, M., Zimmermann, E. *et al*. (2010). Physical activity levels and determinants of change in young adults: a longitudinal panel study. *Int J Behav Nutr Phys Act,* 7(2). https://doi.org/10.1186/1479-5868-7-2

Appendix A

PHYSICAL FITNESS QUESTIONNAIRE

AEROBIC CONDITIONING	RARELY		SOMETIMES		OFTEN
Getting winded from climbing two flights of stairs	1	2	3	4	5
Feeling fatigued or out of breath from taking a brisk walk	1	2	3	4	5
Choosing the elevator over the stairs	1	2	3	4	5
Preferring to drive a few blocks rather than walking	1	2	3	4	5
Making excuses to avoid physical exertion	1	2	3	4	5

TOTAL THE CIRCLED NUMBERS ______

STRENGTH TRAINING	RARELY		SOMETIMES		OFTEN
Difficulty lifting heavy objects	1	2	3	4	5
Concerned about shaking hands with someone who might have a firm grip	1	2	3	4	5
Unable to stand unsupported (e.g., waiting in line) for 15 minutes or longer	1	2	3	4	5
Difficulty opening a window or twisting open a jar top	1	2	3	4	5
Asking others to lift or carry things for you	1	2	3	4	5

TOTAL = ______

BALANCE AND STABILITY	RARELY		SOMETIMES		OFTEN
Losing balance while standing up or sitting down	1	2	3	4	5
Fear of tripping or falling while walking	1	2	3	4	5
Inability to stand on one leg for more than 5 seconds	1	2	3	4	5
Having to sit to put on loafers or slip on shoes	1	2	3	4	5
Feeling unsteady or needing to use a handrail when walking up or down stairs	1	2	3	4	5

TOTAL = ______