RELATIONSHIPS BETWEEN PHYSICAL ACTIVITY, DEMENTIA MORTALITY, AND COGNITION IN FINNISH TWINS

Paula Iso-Markku
The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations

**Supervisors**
Senior Lecturer Katja Waller, M.Sc., Ph.D.
University of Jyväskylä, Faculty of Sport and Health Sciences
Jyväskylä, Finland

Professor Urho M. Kujala, M.D., Ph.D.
University of Jyväskylä, Faculty of Sport and Health Sciences
Jyväskylä, Finland

**Custos**
Professor Jaakko Kaprio
Institute for Molecular Medicine Finland (FIMM)
University of Helsinki, Finland
Department of Public Health
University of Helsinki, Finland

**Reviewers**
Adjunct Professor Kari Kalliokoski
Turku PET Centre
University of Turku, Turku, Finland

Associate Professor Jenni Kulmala
National Institute for Health and Welfare, Helsinki, Finland

**Opponent**
Professor Kaare Christensen
The Danish Aging Research Centre
University of Southern Denmark
Odense, Denmark

ISBN 978-951-51-5503-0 (PDF)

Unigrafia
Helsinki 2019
“A scientist in his laboratory is not a mere technician: he is also a child confronting natural phenomena that impress him as though they were fairy tales”

Marie Curie
ABSTRACT

Physical activity (PA) has been associated with better cognition or decreased dementia incidence in many earlier studies. Since both PA and cognition are, to a large extent, heritable traits, the association found in earlier studies may be confounded by genetic selection. The aim of this thesis study was to ascertain whether midlife PA is associated with late-life dementia mortality and cognition and if objectively measured late-life PA is associated with late-life cognition and whether the possible associations are contributable to genetic factors and childhood shared environment.

The thesis study was implemented in the older Finnish Twin Cohort. Vigorous PA and the volume of PA have been reported in questionnaires in 1975 and 1981. Dementia mortality was followed from the time of the questionnaire in 1981 to the end of the year 2011 (n=21,524). All participants over 65 years of age were offered the possibility to participate in a telephone cognition interview. The cognition interview participants born in 1940–1944 were also offered the possibility to participate in accelerometer monitorings. The participation rates for questionnaires in 1975 and 1981, cognition interviews and accelerometer monitorings were high considering the length of the follow-up (89%, 84%, 78% and 54% among twins alive and with an address in Finland). In the accelerometer study, there was a trend toward a slightly selected population regarding better education and healthier lifestyle.

Among the 21,524 twins followed for dementia mortality, there were 353 dementia deaths. Long-term vigorous PA was associated with decreased dementia mortality (hazard ratio (HR) in the fully adjusted model 0.60, 95% confidence intervals (CI) 0.39 – 0.93). The volume of PA and long-term quantitative physical inactivity were not significantly associated with dementia mortality. Among 3050 twins aged 65 years and older and who had participated in cognition interviews, 204 had cognitive impairment. Long-term vigorous PA was significantly associated with better cognition (beta estimate 0.91, 95% CI 0.47 – 1.35), belonging to the most inactive quintile in both questionnaire years 1975 and 1981 was inversely associated with cognition, and volume of PA was not significantly associated with cognition. Late-life objectively measured light PA was positively and sedentary behavior inversely associated with late-life cognition, but the effect sizes were small. In analyses comparing twins with their co-twins, the point estimates were quite similar to those found at the individual level but were statistically non-significant. The number of twin pairs discordant for PA and dementia mortality or cognition was very small.
Abstract

The results indicate that midlife vigorous activity is significantly associated with decreased dementia mortality and better late-life cognition, but the associations are likely to be explained in part by genetic factors and childhood shared environment. The association of objectively measured PA and cognition in late-life seems weak and, also, explained in part by genetic factors and childhood shared environment.


29 vuotta kestäneen seurannan aikana 21 524 kaksosen joukosta 353 kuoli muistisairauden. Keski-ään pitkääikainenripeä liikunta oli yhteydessä sekä alentuneeseen muistisairauskuolleisuuteen (vakioitu riskitiheysuhde 0,60, luottamusväli [LV] 95% merkitsevyyttä 0,39–0,93) että parempaan kognition yli kaksi vuosikymmenenmäärin terveemmin (vakioitu regressiokerroin 0,91, 95% LV 0,47–1,35). Anos-vasteesuhdetta ei todettu liikunnan määrän ja kognition tai muistisairauskuolleisuuden
välillä, mutta kaksosilla, jotka kuuluivat fyysisesti inaktiivisimpana viidenneksen molempina kyselyvuosina 1975 ja 1981 oli heikompia kognitioja kaksosilla, jotka eivät kuuluneet fyysisesti inaktiivisimpana viidenneksen kumpankaan kyselyvuonna (vakioitu regressiokerroin 0,85, 95% LV 0,22–1,48). Vanhallalle iällä objektiivisesti mitattu passiivinen aika oli käänteisesti ja kevyen liikunnan määrä positiivisesti yhteydessä kognitioon, mutta vaikutus ei ollut suuri (passiivinen aika: vakioitu regressiokerroin – 0,21, 95% LV – 0,42–[0,003], kevyt liikunta: vakioitu regressiokerroin 0,30, 95% LV 0,02–0,58). Parittaisissa analyysissä, joissa vertailtiin kaksosparien sisäisiä eroja liikunnassa kaksosparien sisäisiin eroihin kognitiossa tai muistisairauksesi inäissä, ei löydetyt tilastollisesti merkitseviä yhteyksiä, vaikka trendit olivat saman tyyppisiä kuin yksilöanalyysissä. Sekä liikunnan että muistisairauksesi inäissä kognitioon suhteen diskordantteja eli erilaisia kaksospareja oli erittäin vähän suhteessa mukana olleiden kaksosparien määrään.

Tämän väitöskirjan tulokset osoittavat, että keski-iän ripeä liikunta on yhteydessä parempaan kognitioon ja alentuneeseen muistisairauksesi inäissään yli kaksi vuosikymmentä myöhemmin, mutta yhteys selittyy ainakin osin geneettisillä tekijöillä sekä lapsuus- ja nuoruussi inäissä kasvuynpäristöllä. Objektiivisesti mitatun liikunnan ja passiivisen ajan yhteys kognitioon vanhalla iällä on heikko ja selittyy myös osittain geneettisillä tekijöillä sekä lapsuuden ja nuoruuden kasvuynpäristöllä.
CONTENTS

ABSTRACT ........................................................................................................5
TIIVISTELMÄ ...................................................................................................7
LIST OF ORIGINAL PUBLICATIONS ............................................................ 13
ABBREVIATIONS ...........................................................................................14

1 INTRODUCTION ..........................................................................................16

2 REVIEW OF THE LITERATURE .................................................................18

2.1 Cognition, dementia and dementia mortality ........................................18

2.1.1 Cognition and mild cognitive impairment ........................................18

2.1.2 Dementia ............................................................................................19

2.1.2.1 Alzheimer’s disease ......................................................................21

2.1.2.2 Vascular dementia ......................................................................22

2.1.2.3 Lewy Body Dementia ....................................................................24

2.1.2.4 Frontotemporal dementia ..............................................................25

2.1.2.5 Genetic risk factors for dementia ..................................................25

2.1.2.6 Environmental risk factors for dementia ......................................26

2.1.2.7 Diagnosis and treatment of dementia ..........................................30

2.1.3 Measuring cognition, dementia and dementia mortality ...............31

2.2 Physical activity ......................................................................................34

2.2.1 Definition and prevalence .................................................................34

2.2.2 Heritability of physical activity ..........................................................37

2.2.3 Physical activity’s relationship to cardiorespiratory fitness ......37

2.2.4 Measuring physical activity .................................................................38

2.3 Physical activity and cognition .................................................................39

2.3.1 Physical activity and cognition: prospective cohort studies ....39

2.3.2 Physical activity and cognition: Meta-analyses of prospective cohort studies ..........................................................41

2.3.3 Physical activity and cognition: randomized controlled trials 42

2.3.4 Objectively measured physical activity and cognition ............44
## Contents

2.3.5 Physical activity and dementia: prospective studies .......... 46
2.3.6 Physical activity and dementia: meta-analyses of prospective studies ........................................ 48
2.3.7 Cardiorespiratory fitness and cognition .......................... 51
2.3.8 Resistance training and cognition ................................. 52
2.3.9 Mechanistic studies ...................................................... 53

2.4 Twin studies .................................................................... 57
2.4.1 DNA and genetic selection ............................................. 57
2.4.2 Classical twin design ..................................................... 57
2.4.3 Co-twin control studies ............................................... 58
2.4.4 The assumptions of twin studies ................................. 58

2.5 Summary of the literature review ....................................... 60

3  THE AIMS OF THE STUDY ......................................................... 62

4  PARTICIPANTS AND METHODS ............................................. 63
4.1 Study design and participants ........................................... 63
4.2 Measures ........................................................................ 66
4.2.1 Physical activity ............................................................. 66
4.2.2 Dementia mortality and cognition .................................. 68
4.2.3 Covariates .................................................................... 69
4.3 Ethics of the study ............................................................ 70
4.4 Statistical methods .......................................................... 71

5  RESULTS .................................................................................. 73
5.1 Baseline characteristics .................................................... 73
5.2 Self-reported physical activity and dementia mortality .......... 75
5.3 Self-reported physical activity and cognition after 25-year follow-up ................................................. 81
5.4 Objectively measured physical activity profile and cognition .... 84
6 DISCUSSION........................................................................................................... 86
   6.1 Summary of the main findings................................................................. 86
   6.2 Self-reported physical activity and dementia mortality ...................... 88
   6.3 Self-reported physical activity and cognition ...................................... 89
   6.4 Objectively measured physical activity and sedentary
       behavior profile and cognition .............................................................. 92
   6.5 Strengths and limitations ..................................................................... 95
   6.6 Conclusions ...................................................................................... 100
   6.7 Future prospects ............................................................................ 100

ACKNOWLEDGEMENTS ...................................................................................... 103
REFERENCES ......................................................................................................... 105
APPENDICES ....................................................................................................... 147
ORIGINAL ARTICLES .......................................................................................... 179
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:


These publications have been reprinted with the permission of their copyright holders and are referred to in the text by their Roman numerals.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3MS</td>
<td>The modified Mini-Mental State Examination</td>
</tr>
<tr>
<td>Aβ</td>
<td>β-amyloid peptide</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>AICD</td>
<td>Amyloid precursor protein intracellular domain</td>
</tr>
<tr>
<td>APE</td>
<td>Angle for posture estimation</td>
</tr>
<tr>
<td>Apo E</td>
<td>Apolipoprotein E</td>
</tr>
<tr>
<td>APP</td>
<td>Amyloid precursor protein</td>
</tr>
<tr>
<td>BDNF</td>
<td>Brain-derived neurotrophic factor</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>CADASIL</td>
<td>Cerebral autosomal dominant arteriopathy with subcortical infarct leukoencephalopathy</td>
</tr>
<tr>
<td>CERAD</td>
<td>The Consortium to Establish a Registry for Alzheimer’s Disease</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DSM</td>
<td>The Diagnostic and Statistical Manual of Mental Disorders</td>
</tr>
<tr>
<td>DZ</td>
<td>Dizygotic</td>
</tr>
<tr>
<td>FCAA</td>
<td>Familial cerebral amyloid angiopathy</td>
</tr>
<tr>
<td>FDG</td>
<td>Fluorodeoxyglucose</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association studies</td>
</tr>
<tr>
<td>HDL</td>
<td>High-density lipoproteins</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard ratio</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Diseases</td>
</tr>
<tr>
<td>IGF-1</td>
<td>Insulin-like growth factor</td>
</tr>
<tr>
<td>IL-1β</td>
<td>Intraleukin 1 beta</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligence quotient</td>
</tr>
<tr>
<td>kD</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>LPA</td>
<td>Light physical activity</td>
</tr>
<tr>
<td>LTPA</td>
<td>Leisure-time physical activity</td>
</tr>
<tr>
<td>MAD</td>
<td>Mean amplitude deviation</td>
</tr>
<tr>
<td>MCI</td>
<td>Mild cognitive impairment</td>
</tr>
<tr>
<td>MET</td>
<td>Metabolic equivalent of task</td>
</tr>
<tr>
<td>MID</td>
<td>Multi-infarct dementia</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MMSE</td>
<td>Mini-Mental State Examination</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MVPA</td>
<td>Moderate-to-vigorous physical activity</td>
</tr>
<tr>
<td>MZ</td>
<td>Monozygotic</td>
</tr>
<tr>
<td>NINCDS-ADRDA</td>
<td>National Institute of Neurological and Communicative Disorders and Stroke – and the Alzheimer’s Disease and Related Disorders Association</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>PA</td>
<td>Physical activity</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>QLactive</td>
<td>Active according to long-term quantitative physical activity</td>
</tr>
<tr>
<td>QLchange</td>
<td>Group of change in long-term quantitative physical activity</td>
</tr>
<tr>
<td>QLinactive</td>
<td>Inactive according to long-term quantitative physical activity</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>SB</td>
<td>Sedentary behaviour</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SIVD</td>
<td>Subcortical ischemic vascular dementia</td>
</tr>
<tr>
<td>TELE</td>
<td>Telephone assessment of dementia</td>
</tr>
<tr>
<td>TICS</td>
<td>Telephone Interview of Cognitive Status</td>
</tr>
<tr>
<td>TNF-α</td>
<td>Tumor necrosis factor alpha</td>
</tr>
<tr>
<td>VaD</td>
<td>Vascular dementia</td>
</tr>
<tr>
<td>VLactive</td>
<td>Active according to long-term vigorous physical activity</td>
</tr>
<tr>
<td>VLchange</td>
<td>Group of change in long-term vigorous physical activity</td>
</tr>
<tr>
<td>VLDL</td>
<td>Very-low density lipoproteins</td>
</tr>
<tr>
<td>VLinactive</td>
<td>Inactive according to long-term vigorous physical activity</td>
</tr>
<tr>
<td>VO2</td>
<td>Oxygen consumption</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

The world screams for preventive means to flee from the gaping maw of dementia. The need is personal, societal, and economic. Despite the age-specific risk of dementia declining with younger generations (Langa, 2015), the prevalence of dementia is still expected to rise world-wide because of population aging (Norton et al., 2013). Physical inactivity seems a tempting target for prevention considering its wide prevalence among adults aged 18 years and older: approximately 66% in the U.S. (Fine et al., 2004) and 77% in Finland are inactive according to a recent accelerometer study (Husu et al., 2016). With the new PA recommendations also allowing PA bouts lasting less than 10 minutes to be counted, the numbers may grow smaller (2018 Physical Activity Guidelines Advisory Committee, 2018), but the problem still stares us in the face. Our society runs on sedentariness.

Randomized controlled trials (RCTs) on PA and cognition have thus far been limited by small sample sizes. There are several multi-domain intervention studies, but with conflicting results (Andrieu et al., 2017, Moll van Charante et al., 2016, Ngandu et al., 2015, Lehtisalo et al., 2016, de Souto Barreto, Andrieu et al., 2018). The intervention with the most supervised support in various lifestyle changes did, however, have an effect on cognition (Ngandu et al., 2015), but the contribution of PA alone is unclear.

The majority of prospective studies have started in older years of life or are of short duration considering the long preclinical period of dementia endangering them to misinterpretation because of reverse causation. The amyloid plaques and neurofibrillary tangles might already affect the ability to exercise and an association between PA and declined dementia incidence could be due to preclinical dementia affecting PA engagement. The preclinical phase of dementia can last decades (Bateman et al., 2012, Fagan et al., 2014). The scarce evidence from robust prospective studies of midlife PA and late-life cognition tell a uniform story: PA seems to be beneficial to cognitive aging (Chang, M. et al., 2010, Elwood et al., 2013, Singh-Manoux et al., 2005, Virta et al., 2013a). The long-term follow-up studies from midlife PA and dementia are more controversial: a significant inverse association (Andel et al., 2008, Gelber et al., 2012, Rovio et al., 2005), no significant association (Morgan et al., 2012), significant association only in men (Tolppanen et al., 2015), significant association only for moderate amounts of PA but not high amounts of PA (> 5 hours (h)) (Chang, M. et al., 2010). Earlier, the strong association between PA and mortality had not been refuted but was severely questioned in its strength in a compelling twin study controlling for genetic factors (Karvinen et al., 2015). The possibility that an association between PA and cognition or dementia...
in an observational study is an intermediate effect of body mass index (BMI) also has to be considered. Twin studies on PA and dementia have been scarce and no association between PA and cognition or dementia has been found in analyses comparing twins with their co-twins (Andel et al., 2008, Carlson et al., 2008, Virta et al., 2013a). Does PA protect from dementia and cognitive decline in the absence of confounding factors such as reverse causation and pleiotropic genetic factors?

The aim of this study is to investigate the association of midlife PA with dementia mortality and cognition two decades later, and the association of objectively measured late-life PA and late-life cognition in the older Finnish Twin Cohort. The twin study design allows for co-twin comparisons and thus to take genetic factors and childhood shared environment into account. To my knowledge, this is the largest twin study assessing the association between PA and dementia or cognition to date.
2 REVIEW OF THE LITERATURE

2.1 COGNITION, DEMENTIA AND DEMENTIA MORTALITY

2.1.1 COGNITION AND MILD COGNITIVE IMPAIRMENT

Cognition means “the mental action or process of acquiring knowledge and understanding thought, experience, and the senses,” according to Oxford Dictionaries (Oxford University Press, 2018). Intelligence or general cognitive ability refers to the ability to understand, learn, and solve problems through cognitive processes rather than the actual process called cognition. Healthy cognition is typically perceived as a level of mental functioning that is typical for age. For children, the range of everyday functions that are expected vary rapidly during growth. For young and middle-aged adults, healthy cognition refers to a cognitive level with which one can perform in everyday life independently. In numbers, this is defined as an intelligence quotient (IQ) over 70 (Roivainen, 2015). In older adults, cognitive abilities start to deteriorate even in so-called healthy cognitive aging. The process of healthy cognitive aging is best described briefly with the terms fluid and crystallized intelligence (Cattel, 1971, Horn, 1982). Crystallized intelligence means the ability to use learned skills, knowledge and experience, and this type of intelligence relies on long-term memory. Cognitive processes falling into this category are, for example, vocabulary and semantic memory. Fluid intelligence means the ability to reason, identify complex relationships, and solve new problems. Cognitive processes such as processing speed and working memory fall into this category. Over the life course, fluid intelligence shows decline but crystallized cognitive abilities stay rather stable. The decline in cognitive abilities starts as early as age 45–49 (Singh-Manoux et al., 2012), and its rate is strongly variable between individuals (Wilson, R.S. et al., 2002). This cognitive decline is reflected anatomically in loss of cortical gray matter (Fjell et al., 2010) and in loss of white matter (Guttmann et al., 1998, Ge et al., 2002), even in healthy cognitive aging. The 5th Diagnostic and Statistical Manual of Mental Disorders (DSM-5) divides cognitive abilities into the following six key categories: executive function, perceptual-motor function, language, learning and memory, social cognition and complex attention (Sachdev et al., 2014). These domains can be further divided into subdomains of cognition (Figure 1).
When the rate of cognitive decline exceeds the expected according to age and educational background but does not interfere significantly with activities of daily life, the person suffers from a condition called mild cognitive impairment (MCI) (Gauthier et al., 2006). The majority of these cases progress to dementia within years. Two subtypes are recognized: amnestic and non-amnestic. In amnestic MCI, episodic memory is affected, while in non-amnestic MCI, cognitive domains other than memory such as executive function, language, or visuospatial abilities are affected (Petersen et al., 2014).

2.1.2 DEMENTIA

MCI affects cognitive abilities without interfering with functional abilities (activities of daily living like dressing, eating, walking). When the deterioration of cognitive abilities is so severe that functional abilities are also affected, the condition is called dementia. The official definition from the American Psychiatric Association describes dementia as a syndrome in which memory (ability to learn and retrieve old stored information) and at least one other domain of cognition (language, motor function, perception, executive function) are disturbed so severely that the
person has difficulties in coping independently in everyday life, work, and social relationships (American Psychiatric Association, 2000).

Many different diseases can cause dementia, the most common of which is Alzheimer’s disease (AD), accounting for about 50–70% of dementia cases in both developed and developing countries (Qiu et al., 2007). The second most common cause of dementia is vascular dementia (VaD), accounting for approximately 15–25% of dementia (Qiu et al., 2007). The next two most common neurodegenerative diseases causing dementia are Lewy Body Dementia and frontotemporal dementia, causing respectively about 10–15% (McKeith et al., 1996) and 10% (Hogan et al., 2016) of dementia cases. Other rare causes of dementia are Creutzfeldt–Jakob disease, Human Immunodeficiency Virus, and alcohol. Neuropathology studies, however, tell a different story: rather than attributing cognitive decline and dementia to only one clear causative disease, autopsies often show a complex constellation of underlying pathologies, the number of different pathologies increasing with age (Rahimi et al., 2014).

The global prevalence of dementia in 2010 was estimated to be 35.6 million globally (Prince, M. et al., 2013). According to the most recent estimates, dementia prevalence could be expected to double every 20 years, resulting in a projected figure of 115.4 million dementia cases worldwide in 2050 (Prince, M. et al., 2013). These predictions, however, have been questioned (Norton et al., 2013). Many studies have supported evidence for stable or declining age-specific dementia prevalence (Lobo et al., 2007, Manton et al., 2005, Schrijvers et al., 2012, Rocca et al., 2011, Sheffield et al., 2011, Matthews et al., 2013, Christensen et al., 2013). The speculated reasons for declining age-specific dementia incidence are the higher educational level and more aggressive treatment of hypercholesterolemia and hypertension (Langa, 2015). Contradictions in this area of research still, nevertheless, prevail, while some hypothesize that the driving force of dementia burden derives especially from developing countries, which are inundated with growing vascular disease epidemic (Prince, M. et al., 2016, Gaziano et al., 2010). Due to the population ageing the burden of dementia is, however, going to be substantial regardless of the age-specific dementia incidence.

In Europe alone, the cost of dementia was estimated to be approximately 105.2 billion euros in 2010 (Gustavsson et al., 2011). The worldwide cost is estimated around 742 billion euros (Wimo et al., 2017). While 58% of persons with dementia are living in low- and middle-income countries, the vast majority (86%) of the costs are from high-income countries (Wimo et al., 2017).
2.1.2.1 Alzheimer’s disease

The typical form of AD is characterized first by deficits in episodic memory and difficulties in learning new things. Problem solving, executive function, and the ability to concentrate deteriorate in mild AD. When the disease progresses, the difficulties become more widespread (Erkinjuntti et al., 2007). The patient can suffer from difficulties in language and perception, disorientation, impaired gait and considerable difficulties in episodic memory (Remes et al., 2015). Usually when the disease has progressed to the moderate phase, illness perception weakens (Remes et al., 2015). In severe AD, the symptoms are severe: executive functions are compromised, communication is challenged, perception and practical skills are severely deteriorated (Erkinjuntti et al., 2007). The patient needs a lot of help in every day functions. In atypical forms of AD, the problems with language, visual perception, executive function and practical skills manifest before and are typically more pronounced than memory deficits. This is often the case in familial early-onset dementia (Scheltens et al., 2016).

The neuropathological hallmarks of AD are amyloid plaques and neurofibrillary tangles. Amyloid plaques are aggregated amyloid β (Aβ) situated extracellularly in the brain. Aβ is a peptide produced in the sequential cleavage of amyloid precursor protein (APP). APP is a type I transmembrane protein (Huang, H.C. et al., 2011) whose exact function has remained partly unknown but the protein plays a role in neural stem cell development, neuronal survival, neurite outgrowth and neurorepair (Dawkins et al., 2014). APP is cleaved via two different pathways and only one pathway produces Aβ. Fortunately, the “nonamyloidogenic pathway” (Lu et al., 2003; Nhan et al., 2015; Querfurth et al., 2010) not producing Aβ is predominant (Zhang et al., 2013). An imbalance between the production and clearance of Aβ is considered the culprit in AD pathology. It has been suggested that in familial AD the cause of the imbalance is overproduction of Aβ and that in sporadic AD the problem lies more in defective clearance of Aβ (Blennow et al., 2006). Aβ can be cleared through the blood–brain barrier into the circulation (Shibata et al., 2000), through degradation by proteases and vascular smooth muscle cells, or through lymphatic clearance (Boespflug et al., 2018).

The original “amyloid cascade hypothesis” proposes that abnormal Aβ aggregation initiates the cascade of pathological changes in AD leading to neurodegeneration and cognitive impairment (Hardy et al., 1992). Since the 1990s, the amyloid cascade hypothesis has evolved. The passive accumulation of amyloid plaques is not considered the most detrimental pathological process in AD anymore. Aβ peptides can aggregate to form soluble oligomers. It has been argued that it is the toxic soluble Aβ-derived oligomers that cause synaptic loss, the most robust correlate of AD-related cognitive deficits, by altering the spine density, morphology and composition in dendrites (Lacor et al., 2007, Shankar et al., 2008).
Tau protein is an axonal protein that binds to microtubules, promoting their assembly and stability (Blennow et al., 2006). Its aggregation leads to formation of neurofibrillary tangles (Alonso et al., 1997), another hallmark of AD. The link between these two hallmarks of AD, however, is still elusive (Small et al., 2008) and the exact trigger of the disease process remains a mystery. Neuroinflammation has been recognized to interact widely in the disease progression (Heneka et al., 2015), but its role is unknown; a trigger, a pathway, or a reinforcement, we do not know. The melange of amyloid plaques, Aβ-derived oligomers, neurofibrillary tangles, and neuroinflammation subsequently leads to neuronal degeneration, brain atrophy, and changes in the neurotransmitter system. The brain atrophy in AD starts and is strongest in the hippocampus and in the medial lobes of the brain. It is proposed that the time period from the beginning of the accumulation of neuropathological changes to the onset of clinical symptoms spans from 10 to 30 years (Bateman et al., 2012, Fagan et al., 2014, Jansen et al., 2015, Villemagne et al., 2013).

2.1.2.2 Vascular dementia

VaD is a vast and heterogenous entity regarding both etiology and symptoms. However, the common denominator for the etiology is the vascular origin of the symptoms or the absence or scarcity of AD or Parkinson’s disease pathology. If both vascular pathology and AD pathology co-exist, the disease in question is called mixed dementia. As for the symptoms, the most striking difference to AD is that the function impaired first and more severely, is other than memory deficits (Khan et al., 2016). Typically, deficits in executive function prevail, typically accompanied by motor impairments (O’Brien et al., 2003). Due to the very heterogenous etiologies of VaD, there is, however, a plethora of different clinical presentations. One notable distinction to AD is that mood symptoms are more prevalent (Khan et al., 2016). Mood symptoms comprise anxiety, depression, and emotional lability. Avolition and apathy are typical, particularly in subcortical ischemic vascular disease (Khan et al., 2016).

In short, the following different etiologies and types in VaD have been recognized: multi-infarct dementia, subcortical ischemic vascular dementia (SIVD), strategic-infarct dementia, hypoperfusion dementia, hemorrhagic dementia, and dementia caused by specific arteriopathies (see Figure 2 for summary of the pathophysiological mechanisms in VaD) (O’Brien et al., 2003). Multi-infarct dementia is the form of vascular origin that was first discovered. It manifests with sudden prominent symptoms affecting especially executive functions and motor function. Its progression is stepwise and related to large-vessel occlusions. The form of VaD that was discovered only later is SIVD. This disease is of small-vessel origin and the pathological hallmarks are lacunar infarcts and white-matter lesions (O’Brien et al.,...
Small-vessel disease affects deep parts of the brain in contrast to the multi-infarct dementia affecting especially cortical regions. Lacunar infarcts are caused by atherosclerosis or by lipohyalinosis (Norrving, 2003), a form of segmental arterial disorganization in which the arterial wall is thickened with focal dilation, eventually leading to disintegration of the wall. This causes a small infarct around the site. The process is thought to be inflicted by the pounding pulse from hypertension. Small arteries in the brain are vulnerable because they do not grow smaller gradually, instead they branch straight off large vessels that have thick muscular walls to protect against hypertension (Norrving, 2003). White-matter lesions are areas of high signal intensity on T2-weighted MRI that have been connected with a higher risk of dementia and stroke (Debette et al., 2010) and thought to be indicative of small-vessel disease (Schmidt et al., 2011). However, their exact formation mechanism is not clear, but most likely has to do with small-vessel permeability changes, disruption of the blood–brain barrier, secondary demyelination, and hypoperfusion (Fazekas et al., 2013).

Strategic infarcts occur when a small infarct affects a strategic region in the brain (like the hippocampus) or a large infarct affects subcortical regions (like the brainstem or thalamus) (Khan et al., 2016). Hypoperfusion dementia is associated with local or global hypoperfusion typically associated with vascular risk factors (O’Brien et al., 2003). In hemorrhagic dementia, the damage has been brought on by cerebral hemorrhage. Also, hereditary arteriopathies causing dementia exist: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) and familial cerebral amyloid angiopathy (FCAA).
2.1.2.3 Lewy Body Dementia

Lewy Body Dementia encompasses Parkinson’s disease dementia and Dementia with Lewy bodies (Hanson et al., 2009). Parkinson’s disease is a neurodegenerative disease in which Lewy bodies (i.e., α-synuclein-positive neuronal inclusions) accumulate in the substantia nigra and basal ganglia, causing neuronal damage leading to rigidity of the muscles, rest tremor and bradykinesia. In advanced Parkinson’s disease, cognitive abilities can also be severely affected. Dementia with Lewy bodies is a syndrome characterized by sleep problems, visual hallucinations, cognitive decline and fluctuating confusion. The neuropathological diagnosis of Dementia with Lewy bodies is based on the accumulation of Lewy bodies throughout the cerebral cortex. However, more than half of Dementia with Lewy Body cases have also AD pathology in their brain and it has been suggested that AD, Dementia with Lewy bodies, and Parkinson’s disease represent different aspects of the same disease continuum (Meeus et al., 2012). It must be noted that in Finnish terminology, Lewy Body Dementia denotes solely Dementia with Lewy Bodies while Parkinson’s disease is conceptually a separate entity.
2.1.2.4 Frontotemporal dementia

Frontotemporal dementia is a cluster of neurodegenerative disorders that are characterized clinically by progressive changes in behavior, personality, and/or language (Neary et al., 1998) and anatomically by atrophy of the frontal and temporal lobes. Onset is typically before the age of 65 (Kirshner, 2014). The neuropathological basis lies in the abnormal intracellular accumulation of a protein. The most common proteins to be accumulated are hyperphosphorylated tau, transactive DNA-binding protein of 43kD (kilodaltons) with several abnormal modifications and members in the FET (Fused in liposarcoma, Ewing sarcoma, TATA-binding protein-associated factor 15) protein family.

2.1.2.5 Genetic risk factors for dementia

Heritability refers to the proportion of variance in a phenotypic trait that can be contributed to genetic influences at the population level. It is not stable across generations and different populations from different cultures. In addition to anatomic traits such as hair color or height also skills and preferences in lifestyles are heritable (Kaprio et al., 1987, Laitala et al., 2008). Because heritability is a measure of variance contributed to genetic influences on a population level, the effect of genes in a single individual cannot be determined on the basis of heritability.

The overall heritability of any dementia was estimated to be 43% in a Swedish twin study (Gatz et al., 1997). The heritabilities for different causes of dementia have been estimated to be approximately the following: 33% for AD, < 1% for VaD, 50% for frontotemporal dementia and 40% for Lewy Body Dementia (Ferencz et al., 2015). Additionally, twin studies have also shown a remarkably large heritability for general cognitive ability, with genes accounting for 50% of the variance in general cognitive abilities across 11,000 twin pairs (Haworth et al., 2009). Genome-wide association (GWA) studies have even been able to identify genetic variants that account for 20% of the 50% heritability of general cognitive ability (Plomin et al., 2018).

AD can be divided into two different types: early-onset dementia and late-onset dementia. The diagnostic cut-off value is 65 years. Early-onset dementia is typically familial caused by gene mutations in APP, PSEN1, or PSEN2, although sporadic forms also occur. APP, PSEN1, and PSEN2 gene mutations are rare, autosomal, and dominant with nearly 100% penetrance (Rocchi et al., 2003). They are all part of the production of Aβ; APP being the precursor protein from which Aβ is cleaved from and PSEN1 and PSEN2 being subunits of the cleavage enzyme. The vast majority, approximately 95%, of AD is caused by the late-onset AD type (Rocchi et al., 2003), which is typically sporadic although genetic susceptibility to late-onset
AD exists. Genetic susceptibility for late-onset AD is conferred by numerous quite common genetic risk factors with low penetrance and by one moderately penetrant gene (APOE e4). The overall heritability in late-onset AD is estimated to be around 58–79% based on a Swedish twin study (Gatz et al., 2006). In Finnish twins, the pairwise concordance for AD in MZ twins was 18.6% and in DZ twins 4.7% (Rääihä et al., 1996), also indicating a genetic component. The APOE gene codes for apolipoprotein E, which mediates cholesterol metabolism (Mahley, 1988) and has many other functions as well (Yu, J.T. et al., 2014). The low penetrance risk genes linked to an increased risk of AD through GWA studies are implicated in immune response, endocytosis, synaptic function, and cholesterol metabolism (Karch et al., 2015). The single nucleotide polymorphisms found in GWA studies explain approximately 10% of the heritability of AD (Brainstorm Consortium et al., 2018).

As mentioned above, the heritability of VaD is estimated to be <1% (Ferencz et al., 2015). However, there are two highly heritable forms of VaD: CADASIL and FCAA. Notwithstanding these rare forms of VaD, the heritability for VaD seems to be very low. Only one twin study has attempted to elucidate the heritability of VaD (Bergem et al., 1997). They found no significant genetic component. Later, researchers identified some candidate genes for VaD but these findings have not been replicated (Ferencz et al., 2015). The heterogeneity of the etiology of VaD complicates this field of research.

Despite the heterogeneous nature of frontotemporal dementia, this disorder cluster has been found to be highly heritable (Po et al., 2014). Heritability does vary within frontotemporal dementia between the different syndromes (Rohrer et al., 2009), and the forms of frontotemporal dementia called the behavioral variant frontotemporal dementia and frontotemporal dementia with amyotrophic lateral sclerosis are suggested to drive this high heritability (Po et al., 2014).

Different approaches to quantifying the heritability of Dementia with Lewy bodies have yielded different results. The twin study approach has not supported a genetic etiology for Dementia with Lewy bodies (Wang, C.S. et al., 2009), but Dementia with Lewy bodies does aggregate in families (Nervi et al., 2011) and GWA studies suggest a heritability of 36% (Guerreiro et al., 2018).

2.1.2.6 Environmental risk factors for dementia

Environmental factors are major contributors in the development of dementia, although the influence of genetic factors is also substantial. For example, according to a recent estimate, about one-third of dementia cases due to AD, the most common cause of dementia, can be attributed to seven modifiable risk factors, and by reducing each risk factor by 10%, the prevalence of AD could be reduced by 8% worldwide (Norton et al., 2014). These seven risk factors are diabetes, midlife hypertension,
midlife obesity, physical inactivity, depression, smoking, and low educational attainment (Norton et al., 2014). It must be noted, however, that studying causal connections for dementia is challenging because of the decades-long development process of dementia. RCTs are essentially impossible to perform and cohort studies are liable for confounding. The evidence presented in this chapter is mainly based on cohort studies.

Diabetes has been linked to increased incidence of dementia in many longitudinal studies (Cheng et al., 2012) and through several pathways: the vascular pathway, glucose toxicity, and insulin resistance. Diabetes increases strokes in the brain and causes microvascular damage (Biessels et al., 2006). High glucose levels induce abnormal protein glycation and oxidative stress, damaging the normal functions of a brain cell (Biessels et al., 2006). High insulin levels are linked to higher rates of β-amyloid peptide (Aβ) secretion and reduced Aβ clearance (Biessels et al., 2006). The most significant negative impact seems to be the increase of cerebrovascular pathology (Vagelatos et al., 2013, Ahtiluoto et al., 2010). Common genetic etiology in type 2 diabetes and AD has also been recognized in GWA studies (Hao et al., 2015).

The relationship of blood pressure and dementia has been found to be age-dependent. In middle age, hypertension is associated with an increased risk of dementia, while in the elderly very high blood pressures (e.g., systolic blood pressure > 180 mmHg) and low diastolic blood pressure are especially detrimental (Qiu et al., 2005). Whether the association of blood pressure and dementia is causal is unclear. Blood pressure and dementia may also both be affected by a common denominator like a genetic factor, education, or other lifestyle component (Qiu et al., 2005).

The relationship of body weight and dementia is not linear either. Both midlife underweight and midlife overweight or obesity have been associated with an increased risk for dementia (Anstey et al., 2011, Singh-Manoux et al., 2018), but in later life, it is weight loss that preceeds dementia (Barrett-Connor et al., 1996, Singh-Manoux et al., 2018, Kivimäki et al., 2018). However, in a Finnish twin study, the correlation between midlife BMI and old age cognition was mostly explained by genetics (Lahtela et al., 2011). On the other hand, an animal study has linked obesity to the disruption of blood–brain barrier, neuroinflammation, and oxidative stress (Tucsek et al., 2014).

The cognitive reserve hypothesis states that individuals with better cognitive reserve can tolerate more neuropathological changes without developing symptoms of a memory disorder compared to those without cognitive reserve (Meng et al., 2012, Stern et al., 2018). In practice, a person’s educational and occupational attainment are thought to reflect their cognitive reserve. Cognitive reserve is a particularly weighty factor in dementia prevention as the summary odds ratio (OR) against dementia in a comprehensive meta-analysis was as low as 0.54 with narrow confidence intervals (CI) (95% CI 0.49–0.59) (Valenzuela et al., 2006). In addition, a simple risk score based on self-report of midlife educational and
occupational attainment has been found to predict dementia risk well even 20–40 years later (Vuoksimaa et al., 2016). How does this mystery tool against deteriorating brain function work? The current conception is that the resilience of the brain is explained both quantitatively, that is in the sheer brain volume or in the amount of synapses or neurons in the brain, and qualitatively (Stern, 2012, Stern et al., 2018). A brain with high cognitive reserve is thought to use pre-existing networks in a more efficient way in the face of pathology and also, to recruit compensatory networks more efficiently (Stern, 2012, Stern et al., 2018). The brain is, thus, more resilient and plastic to cope with problems.

According to meta-analyses, a history of depression may increase the risk of AD and late-life depression is associated with an increased risk for all-cause dementia (Ownby et al., 2006, Diniz et al., 2013). Multiple mechanisms have been speculated, including shared vascular risk factors, possible genetic links, and low-grade inflammation (Ownby et al., 2006). Interestingly, recent research has suggested low brain-derived neurotrophic factor (BDNF) levels to be a common denominator in depression and dementia (Weinstein et al., 2014). Late-life depression can, however, also be a symptom of preclinical dementia just like reduced late-life physical activity can be a sign of early dementia instead of being a risk factor, as I discuss later in this literature review.

Many other conditions and lifestyle factors have also been considered. Heavy smoking and head injuries have been considered risk factors for dementia according to recent meta-analyses (Perry et al., 2016, Xu et al., 2015). Midlife hypercholesterolemia is possibly and probably involved in the development of dementia and AD according to current evidence (Kivipelto et al., 2006, Power et al., 2018). In late life, however, a pattern of serum cholesterol lowering is recognized before the diagnosis of dementia (Kivipelto et al., 2006, Anstey et al., 2008). The nature of the relationship between hypercholesterolemia and dementia is also unclear. APOE ε -status or other genetic factors may affect both. Physical activity and inactivity are discussed later in this literature review. Both heavy users of alcohol and abstainers from alcohol are probably at greater risk for dementia (Sabia et al., 2018, Virta et al., 2010). Feelings of loneliness (Holwerda et al., 2014), midlife stress (Andel et al., 2012, Crowe et al., 2007, Johansson et al., 2010, Sindi et al., 2017, Wang, H.X. et al., 2012), and both short and extended sleep (Bokenberger et al., 2017, Virta et al., 2013b, Wennberg et al., 2017) have been associated with increased risk of dementia. Among non-modifiable risk factors, age is the most significant risk factor for dementia. A summary of the risk factors of dementia is presented in Figure 3.
Figure 3. Risk factors for cognitive decline. According to Norton et al. (2014), about a third of the world’s dementia cases are attributable to midlife obesity, midlife hypertension, diabetes, physical inactivity, low educational attainment, smoking, and depression. The evidence for many of these risk factors including physical inactivity with cognitive decline is, however, indecisive due to the lack of long-term follow-ups and good quality RCTs. It is also not clear what kind of PA is ideal to protect from cognitive decline. (Level of evidence: + low, ++ moderate, +++ robust) (Prince, M. et al., 2014, Plassman et al, 2010, Erkinjuntti et al., 2007)
2.1.2.7 Diagnosis and treatment of dementia

The definite diagnoses of dementias are based on neuropathological examination post mortem. Clinical diagnosis of dementia can, at best, be probable, but with new positron emission tomography (PET) and cerebrospinal fluid (CSF) markers the diagnosis of AD is more specific.

For the diagnosis, several steps have to be taken. A doctor’s face-to-face clinical evaluation is needed (Scheltens et al., 2016). The symptoms and a detailed history are taken both from the patient and from a person near to the patient like a close relative (Rinne et al., 2010). Extensive neuropsychological evaluation and assessment of activities in daily living have to be done. Blood tests are used to exclude other causes of dementia or to identify co-morbidity or risk factors. Imaging is a crucial part of the clinical evaluation. Magnetic resonance imaging (MRI) is the gold standard. Computer tomography (CT) is used if MRI is not applicable. In MRI, tumours, hydrocephalus, and subdural hematoma are ruled out and the distribution of brain atrophy is evaluated. As mentioned above, CSF markers, 2-deoxy-2-(18 F) fluoro-D-glucose (FDG) PET/CT, or PET/CT with an amyloid ligand can be used for further precision in the diagnosis. In FDG PET/CT, AD is characterized by diminished glucose metabolism in parieto-temporal areas and in the posterior cingulate cortex, while in frontotemporal dementia diminished glucose metabolism is most pronounced in the temporal and frontal lobes. The novel PET tracers enable the imaging of amyloid plaques in vivo. In CSF markers, decreased Aβ42, elevated total-tau, and phosphorylated tau refer to AD (Olsson et al., 2016). Tau imaging with PET has also become possible in recent years, but in vivo studies with the more specific second-generation tau ligands are still scarce (Leuzy et al., 2019).

No cure for any of the neurodegenerative disorders exists. The treatment of dementia consists, nevertheless, of many different modalities. The most important part of the treatment is supportive care for the patient and his or her family (Erkinjuntti et al., 2007), including arrangements like home care or housing, a guardian of interests, a living will, management of concomitant medical conditions or polypharmacy aggravating the situation, and driving license revocation. Social visits and interaction are very important for the patient. Cognitive training may be beneficial at early stages of AD (Yu, F. et al., 2009) and exercise has also proven to have a positive effect on cognition in patients with MCI (Öhman et al., 2014) and perhaps even in patients with a full-blown dementia (Groot et al., 2016).

Symptomatic medications exist for AD. Cholinesterase inhibitors (donepezil, rivastigmin, galantamine) and memantine improve cognition and stabilize function for a period of time (Birks et al., 2006, Birks et al., 2009, Loy et al., 2006, McShane et al., 2006). Some cholinesterase inhibitors may also improve cognition in VaD, Parkinson’s disease dementia, and Dementia with Lewy bodies, but the level of evidence is more scarce (Suomalaisen lääkäriseuran Duodecimin, Societas
For frontotemporal dementia, no symptomatic medication exists. For AD, many disease-modifying treatments have been investigated. Immunotherapy targeting Aβ (Ballard et al., 2011), influencing the cleavage enzymes of APP (Ballard et al., 2011), anti-aggregation drugs (Kumar et al., 2015), and therapies targeting tau phosphorylation (Kumar et al., 2015) have all failed thus far. A promising branch of disease-modifying medication research is stem cell research, but success in human trials is still awaited (Duncan et al., 2017).

2.1.3 MEASURING COGNITION, DEMENTIA AND DEMENTIA MORTALITY

When studying the incidence of cognitive decline irrespective of the severity, be it dementia also affecting functional abilities or not, the researcher aims to find all possible cases of cognitive decline. However, the financial resources may set restrictions. Basically, the researcher can measure cognition with a measuring instrument, study hospital discharge registers and Drug Reimbursement Registers or resort to the Cause of Death Registers. Figure 4 illustrates how efficient these three methods generally are in studying cognitive decline. Depending on the instrument, measuring cognition is most likely to detect cognitive decline best and to cover the full range of it: from MCI to severe dementia. When resorting to dementia diagnoses from registers (e.g., hospital discharge registers and Drug Reimbursement Registers), the researcher is most likely to find cognitive decline at the level of dementia. However, with this measure, the researcher is likely to miss MCI and some patients with dementia who have been able to stay at home with their dementia or dementia that was not been recorded for some reason in the hospital discharge diagnoses. It must be noted that there is no medication for frontotemporal dementia, thus, this cause of dementia is likely to be missed in Drug Reimbursement Registers. The tip of the iceberg, detecting only the most severe dementia cases, is measuring dementia mortality.

One can measure cognition in face-to-face interviews, with instrumented or computer-based administration methods, or via telephone interview. The best measure of cognition depends on the age and educational background of the participant or cohort in question. The Mini-Mental State Examination (MMSE) (Folstein et al., 1983), CERAD (The Consortium to Establish a Registry for Alzheimer’s Disease) (Chandler et al., 2005), and 7 Minute Screen (Solomon, P.R. et al., 1998) are cognitive tests meant to be used to screen dementia in older adults. They all test several different domains of cognition and CERAD is the broadest but also the most time-consuming. The MMSE is the most commonly used and
it tests for orientation, registration, attention, calculation, and recall. These are administered in face-to-face interviews with paper and pen. It must be taken into account that fatigue, acute illnesses, medication use (benzodiazepines, psychotics), and distractions may affect the evaluation. One of the limitations of MMSE is that it poorly identifies cognitive decline in highly educated individuals.

Cognitive tests applied via telephone have also been developed. In research, they enable the overcoming of obstacles such as geographical distances and might even improve follow-up rates and prevent sample attrition because of its easiness to participate (Kliegel et al., 2007). Telephone-administered cognition screening tools are not to be used in diagnosing dementia, but can be useful as screening instruments (Järvenpää et al., 2002, Kliegel et al., 2007, Castanho et al., 2014, Manly et al., 2011). The evident limitation for telephone-administered cognition screening tools is hearing problems that might distort the assessment (Roccaforte et al., 1992). Verifying hearing is, thus, important. When used correctly, telephone-administered cognition screening tools have been proven to be as reliable and valid as information obtained in face-to-face interaction (Kliegel et al., 2007, Wilson et al., 2005, Järvenpää et al., 2002). No one telephone-based cognition screening instrument has proven to be preeminent (Martin-Khan et al., 2010). The telephone assessment of dementia (TELE) and Telephone Interview of Cognitive Status (TICS) are examples of telephone-based cognition screening instruments. They are reliable and valid tools for cognition screening in elderly populations (Gatz et al., 2002, Castanho et al., 2014, Brandt et al., 1998). They both address the lower end of cognition and dementia, but assessment of MCI is not validated (Castanho et al., 2014, Kliegel et al., 2007). TICS assesses orientation to time and place, attention, short-term memory, sentence repetition, immediate recall, naming to verbal description, word opposites, and praxis, while TELE addresses activities of daily living, visuospatial function, attention memory, long-term memory, calculation, and language. Both instruments have been validated in the Finnish population and correlate well with the results of the MMSE (Järvenpää et al., 2002). TICS is also the most widely translated and validated telephone cognition screening instrument (Castanho et al., 2014).

National registries in Finland are in general considered very accurate. In 2014, a validity study on dementia and AD diagnoses in Finnish national registries was published (Solomon, A. et al., 2014). The sensitivity and positive prediction value to detect dementia from the Hospital Discharge Registry and the Drug Reimbursement Registry together was 71.1% with a positive predictive value of 100%. The Hospital Discharge Registry alone had a poor sensitivity to detect dementia cases, 17.7%, and its positive predictive value was specially limited before 1998. The authors suggest that major modifications in the International Classification of Diseases (ICD) criteria, increasing dementia awareness and the introduction of several new drugs
to dementia are likely causes of the improvement in the positive predictive value. In England, the sensitivity and specificity of dementia recording according to hospital records were 78% and 92%, respectively (Sommerlad et al., 2018). Sommerland et al. (2018) also found that the accuracy of recording of dementia diagnoses is lower especially for people from ethnic minorities, younger, single people, and those with physical illnesses. In Sweden, very different estimates of sensitivity to detect dementia in hospital discharge registries have been found (26% (Dahl et al., 2007) or 82% (Feldman et al., 2012), according to the different gold standard of dementia used in the studies (consensus based on MMSE result, medical records and hospital discharge registry (Dahl et al., 2007) or information based on telephone screening of dementia in conjunction with an academic study (Feldman et al., 2012)). A review of the accuracy of dementia diagnoses according to routinely collected health data also reported large heterogeneity in accuracy estimates (Wilkinson et al., 2018).

The cause of death is a disease or event that led to death. A doctor marks down the cause of death on a death certificate defining the underlying cause of death, acute cause of death and a maximum three contributing causes of death. According to the World Health Organization’s definition, the underlying cause of death means “the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury” (WHO, 2018). The acute cause of death is the acute disease or injury from which a person dies. Contributing causes of death mean the diseases, circumstances, or injuries that contributed to the cause of death. It is clear that when a registry-based indicator is used as a measure of dementia incidence, all possible dementia cases are not found, only the most severe ones. It is, however, likely that moderate and severe dementia, in which a person has difficulties in independent living, is recorded as the underlying or contributing cause of death because it is medically correct because of the weakening and disabling nature of the disease even if the acute cause of death is related to, for example, coronary artery disease. The severe disease seldom goes unnoticed because of the loss of the ability to live independently. According to a validity study, the sensitivity of the Causes of Death Register to detect dementia cases in Finland is 62.2% with a positive predictive values of 100% (Solomon et al., 2014). Solomon et al. (2014) deduced that an underestimation in dementia diagnoses from registries may lead to an underestimation of association with risk or protective factors. On the other hand, a Swedish validity study reported that the Swedish Causes of Death Registry had a sensitivity of 90% to detect dementia cases (Feldman et al., 2012).

Competing causes of death and changes in their treatment or incidence rates affect the incidence rates of dementia and dementia mortality. Because the incidence and mortality of major cardiovascular diseases have significantly decreased since the 1980s (Di Cesare et al., 2013, Nichols et al., 2013, Rosengren et al., 2013), the persons with the heaviest burden of vascular disease factors live longer. Vascular risk
factors are also implicated in dementia, hence the significant decrease in mortality of cardiovascular disease may be one reason leading to higher dementia prevalence.

![Figure 4](image_url)

**Figure 4.** Measuring cognition, dementia, and dementia mortality in research. Adapted from free-hand sketch from Professor Urho Kujala, with his permission.

### 2.2 Physical Activity

#### 2.2.1 Definition and Prevalence

According to the World Health Organization, PA is defined as “any bodily movement produced by skeletal muscles that requires energy expenditure” (WHO, 2010). Some definitions further specify that the energy expenditure has to be “substantial” (Howley, 2001). PA has multiple remarkable health benefits (Lee, I.M. et al., 2012), while the influence of physical inactivity on health is especially considerable because the prevalence of physical inactivity is estimated to be very high, about two-thirds (66%) globally (Fine et al., 2004).

What amount and intensity of exercise is beneficial for health? The answer is simple: all PA is beneficial, although adverse effects increase when very high amounts of exercise are reached (Eijsvogels et al., 2018). The 2018 Physical Activity Guidelines for Americans recommend at least 150 minutes of moderate-intensity PA or at least 75 minutes of vigorous-intensity PA per week for substantial health benefits (2018 Physical Activity Guidelines Advisory Committee, 2018). The new recommendation does not require the PA bouts to last at least 10 minutes, as did the older recommendation. Muscle strengthening is recommended at least two times per week and preferably to all major muscle groups. The issue of the intensity of PA is complex. While it is straightforward to set limits on absolute intensity of
PA, the relative intensity is very variable among individuals of different fitness levels. The general definition of absolute exercise intensity is described in energy expenditure: light exercise meaning 1.1–2.9 METs (metabolic equivalent of task, i.e., multiples of metabolic resting energy expenditure), moderate exercise 3.0–5.9 METs, and vigorous exercise 6.0 or more METs (The Physical Activity Guidelines Advisory Committee, 2009). However, the relative intensity of, for example, brisk walking (3.8 METs) can be very light for a fit person and vigorous for a non-fit person (see Figure 5) (Howley, 2001, Kujala et al., 2017). This variability in perceived intensity levels also makes research and categorization of physical intensity levels more challenging.


During the past decade, new research data has shown that not only is long-lasting PA important to health but that an active lifestyle doing something else than sitting for long periods of time is also beneficial regarding health. In fact, sedentary time has been associated with greater mortality risk independent of moderate to vigorous PA (Koster et al., 2012). Sedentary time has also been associated with greater incidence of cardiovascular disease, cancer and type 2 diabetes, again independent of PA
(Biswas et al., 2015). Contradictory findings exist however. In a large meta-analysis, high levels of moderate intensity PA eliminated the increased risk for mortality associated with high sitting times (Ekelund et al., 2016). A recent cross-sectional study showed that PA was associated with a better cardio-metabolic risk profile even with high concomitant sedentary times (Bakrania et al., 2016). What seems clear though is that both sedentary time and lack of moderate PA confer health risks, even if being proficiently active in one might lessen the inactivity risk of the other.

Interesting evidence has also been provided by studies comparing the health associations of sedentary time used in different contexts. Depending on what the sedentary time is used for: for occupational aims or more passively watching TV for example, recent studies have found distinct associations with health. The evidence indicates that occupational sitting time is less deleterious for cardiometabolic risk factors and cardiorespiratory fitness than leisure sitting time (Saidj et al., 2013, Saidj et al., 2014, Pinto Pereira et al., 2012).

The amount of sedentary time in the United States has been estimated in an accelerometer study to be around 55% (7.7 hours/day) for all participants, both children and adults, and for older adolescents and adults 60 years or over as high as 60% of their waking hours (Matthews, C.E. et al., 2008). In Finland, the percentage of waking hours spent sedentary has been estimated to be 59%, the daily time spent on light PA, on average, 2 hours (h) 8 minutes (min), and the daily time spent in moderate- to vigorous PA, on average, 1 h 17 min (Husu et al., 2016). Only 22.5% of the sample members fulfilled the previously-mentioned old weekly recommendation for health-enhancing aerobic PA requiring the PA bouts to last at least 10 min (Husu et al., 2016). This number is significantly smaller than the one retrieved from the questionnaire studies. According to questionnaire studies, 50% of working-age adults fulfilled the aerobic exercise criteria for health recommendations, but only 18% among men and 16% among women fulfilled the recommendations for muscle-strengthening PA (Husu et al., 2011). These studies were implemented when PA recommendations of that time recommended at least 150 min of moderate PA or at least 75 min of vigorous PA every week consisting of PA bouts lasting at least 10 min. If PA bouts longer than 1 min are included in the calculation like the new guidelines recommend (2018 Physical Activity Guidelines Advisory Committee, 2018), the aerobic PA levels are much higher than if only bouts longer than 10 minutes are included. For example, in a cross-sectional study of over 9000 Finnish employees, approximately 80% of men and 56% of women fulfilled the aerobic PA recommendation measured with an accelerometer (Mutikainen et al., 2014).
2.2.2 HERITABILITY OF PHYSICAL ACTIVITY

In a twin study combining data from seven different European countries, the heritability of leisure-time PA (LTPA) in general has been estimated to be around 48–71% (Stubbe et al., 2006). The effect genes have on PA participation seems, thus, to be of moderate size. The effect of genetic influences on participation in vigorous PA and more intense sports has been estimated to be larger than the genetic influences on lighter and less intense PA (Beunen et al., 1999, Mustelin et al., 2012, de Geus et al., 2014). In a Finnish twin study, the heritability of LTPA was estimated at different ages and the heritability was found to be moderate (43–52%) in adolescence and declined to about 30% in young adulthood (Aaltonen et al., 2013). The estimates were similar for both men and women. These studies have relied on self-reports of PA in estimating PA engagement. The arrival of practical accelerometers and cheap heart rate monitors on the markets has enabled the estimation of heritability of PA with objective measures independent of recall bias or social desirability bias. Heritability of objectively measured moderate-to-vigorous PA (MVPA) has been estimated to be around 47% (den Hoed et al., 2013) and heritability of sedentary behavior around 35–41% (den Hoed et al., 2013, Waller et al., 2018). A recently published GWAS from the UK identified ten genetic variants associated with engagement in habitual PA and was the first study to do so (Klimentidis et al., 2018). Among these were genetic variants in CADM2, a gene previously linked to BMI variation and in APOE, a gene strongly implicated in AD.

2.2.3 PHYSICAL ACTIVITY’S RELATIONSHIP TO CARDIORESPIRATORY FITNESS

Cardiorespiratory fitness means “the ability of the circulatory and respiratory systems to supply oxygen during sustained physical activity” (WHO, 2010). It is usually expressed as maximal oxygen uptake and can be measured in maximal exercise testing procedures with gas analyses (Shepard et al., 1968). The heritability of cardiorespiratory fitness has been estimated to be about 60–70% according to twin studies (Fagard et al., 1991, Sundet et al., 1994) and about 40–50% in family studies (Montoye et al., 1978). However, no single genes affecting cardiorespiratory fitness with certainty are known (Sarzynski et al., 2016). Regular exercise leads to adaptive changes in the cardiorespiratory system which improve fitness, but there are great differences in the responsiveness of maximal oxygen uptake to training and these differences are explained to a great extent by shared familial factors, including genetics (Zadro et al., 2017, Bouchard et al., 2001, Mustelin et al., 2011). However, a recent study showed that such a phenomenon as “non-responders” does not exist. By increasing the amount of exercise, the individuals with no response in cardiorespiratory fitness to lower amounts of endurance training also begin to show improvements (Montero et al., 2017).
2.2.4 MEASURING PHYSICAL ACTIVITY

The most commonly used method for measuring PA in the past has been through questionnaires. The use of self-reports like this has been cost-effective and easy to apply to large cohorts. The quality of PA measurement also depends on the questionnaire used to estimate PA. Structured detailed questions covering the intensity, duration, and frequency of PA yield more reliable results than crude vague questions with yes or no answer options. However, high-quality PA questionnaires also have limitations, like recall bias, biased results due to overestimating PA for social desirability, and both cultural and individual differences in perception (Tucker et al., 2011, Warren et al., 2010). For example, in a considerable cohort of US adults, the estimates of the proportion of adults meeting the health-enhancing PA recommendations varied greatly between data from self-reports and data from objective measurement using accelerometer data (Tucker et al., 2011). In this study, the proportion of adults meeting the recommendations was approximately 59.6% according to self-reports, the corresponding figure from objectively measured data was only 8.2%. Another study systematically reviewed objective measurement and self-report of PA and found both higher and lower levels of PA in self-reports compared to direct data (Prince, S. A. et al., 2008). Thus, comparison between studies using objective or self-report measures of PA is very difficult.

Measuring PA with an accelerometer has, nevertheless, its own weaknesses. Many activities rather common in Finnish society cannot be measured accurately with a simple single accelerometer, like the works of a lumberjack, bicycling, swimming, water running or water walking, and other aquatic exercise. Lifting heavy weights or working out at a gym are also outside of the method’s accurate measurement capability. Additionally, the method does not differentiate between different types of ground – it is far more strenuous to walk in a swamp or on very rough ground with many obstacles like logging waste than on even ground. There are also differences between accelerometers. In general, the hip-worn accelerometers are considered more reliable because of their ability to differentiate between standing still, sitting, and lying (Rosenberger et al., 2013). One further limitation in comparison between PA studies using the accelerometers is the high variability in the analysis and interpretation methods (Kowalski et al., 2012). A Finnish study group has, however, proposed a uniform interpretation method for accelerometer data irrespective of the brand: the mean amplitude deviation with universal cut-off limits for intensity, providing better comparability (Vähä-Ypyä et al., 2015a). The optimal way to measure PA would probably be to combine accelerometer and heart rate monitoring with a preceding laboratory exercise test to determine individual PA intensity levels (Sievänen et al., 2017). The accelerometer would accurately cover the amount of sedentary behaviour (SB) and light PA (LPA), and the heart rate monitor would accurately differentiate between more strenuous PA with the help of individually set threshold levels.
2.3 PHYSICAL ACTIVITY AND COGNITION

2.3.1 PHYSICAL ACTIVITY AND COGNITION: PROSPECTIVE COHORT STUDIES

Numerous prospective cohort studies addressing PA and cognition or dementia exist (see Appendix I). The majority of these studies did not differentiate between aerobic exercise and resistance training, thus the term ‘physical activity’ is not specific to either aerobic exercise or resistance training but can encompass both training types. I have tried to find all studies assessing PA and cognition in adulthood or in older age, but I have not done a systematic review. I do, however, assess their quality and results in a narrative manner in this chapter. The studies are vastly heterogenous. The majority of them have assessed PA with a questionnaire and only a few with an objective measure of PA. Although the questionnaire method unifies the majority of studies, the questionnaires have been varied and the categorization of PA has been heterogenous. This makes the studies difficult to compare. Furthermore, the majority of the studies have been rather short in duration and have been conducted in an elderly population. This kind of setting makes the study vulnerable to bias. Considering the lengthy preclinical phase in AD (Fagan et al., 2014, Jansen et al., 2015, Vilmagne et al., 2013, Bateman et al., 2012), physical inactivity might reflect a preclinical state of dementia in a study with short follow-up. An ideal setting for studying the effect of PA on cognition would be a RCT lasting for decades. Obviously, this kind of study would be too expensive to conduct, the drop-out rate would become high, and not to mention it would be unethical for the physical inactivity group because of all the benefits of PA. Consequently, RCTs can only do so much to prove causality in quite short-time periods. Epidemiological studies are needed to reveal associations in risk or protective factors for diseases whose development lasts years or decades. When PA has been measured in midlife, it is unlikely that the participants suffer from preclinical dementia. On this account, epidemiological studies lasting long and starting in midlife are important.

The number of prospective cohort studies addressing PA and cognition that have started in the participants’ midlife and with a duration of over 10 years is, to my knowledge, 8 (Richards et al., 2003, Sabia et al., 2017, Elwood et al., 2013, Gross et al., 2017, Chang, M. et al., 2010, Virta et al., 2013a, Kåreholt et al., 2011, Singh-Manou et al., 2005). The mean age at the time of PA assessment ranged from 36 to 57 years in the studies that reported mean age in addition to age range. Follow-up length varied from 11 to 30 years. PA was measured heterogeneously (number of different physical activities vs. volume of PA and participation in vigorous PA). Nearly all studies adjusted for age (one study did not, but their cohort was all the same age (Richards et al., 2003)) and sex (except for one studying only men (Elwood et al., 2013)). Six studies controlled for education while two studies lacking adjustment for education adjusted for socio-economic status (Sabia et al., 2017) or social class and a test measuring intelligence in mid-follow-up (Elwood et al., 2013).
One study did not adjust for education, social class, or a measure of crystallized intelligence, but their cohort consisted of medical graduates, thus, having a very similar educational background (Gross et al., 2017). Three studies did not take into account vascular risk factors other than PA (Elwood et al., 2013, Sabia et al., 2017, Richards et al., 2003) and two studies also controlled for a measure of intelligence, either intelligence quotient (Richards et al., 2003) or a test describing crystallized intelligence (Singh-Manoux et al., 2005). The majority of studies found an inverse association between PA and some measure of cognition (Chang, M. et al., 2010, Elwood et al., 2013, Virta et al., 2013a, Singh-Manoux et al., 2005). Kåreholt et al. (2011) found an association between PA and items from the MMSE only in women, while Richards et al. (2003), Sabia et al. (2017) and Cross et al. (2017) did not find a significant association between PA and cognition.

Overall, four of these studies were of high quality, providing sound measures of both PA and cognition, controlling for the most important confounding factors (age, sex, education or premorbid intelligence), and having at least a moderately large unselected study cohort (>1000) (Singh-Manoux et al., 2005, Chang, M. et al., 2010, Virta et al., 2013a, Elwood et al., 2013). The result in all of these was in favor of PA: PA was associated with better cognition either globally or in some domain of cognitive function. In the study of Singh-Manoux et al. (2005), PA was associated with only fluid intelligence but not memory or phonemic or semantic fluency. By contrast, in the study of Chang et al. (2010), all three areas of cognition measured (processing speed, executive function, and memory) were statistically significantly associated with PA. In the study of Virta et al. (2013a), the lowest quartile of PA had a significantly higher risk of cognitive decline than the highest quartile of PA. Elwood et al. (2013) reported an OR of 0.62 with 95% CIs of 0.41–0.92 for regular exercisers developing cognitive decline. For studies with measures of PA that were cruder, the results were different: in the study of Kåreholt et al. (2011), midlife participation in sports, gardening, and dancing was associated with cognition only in women, and in the study of Richards et al. (2003) the number of midlife physical activities did not have an independent association with verbal memory. Therefore, robust and methodologically sound studies examining midlife PA and late-life cognition are scarce, but in favor of a positive association.

While the majority of observational studies of PA and cognition in the elderly are of quite short duration, a small number with a study duration exceeding seven years exists (Yaffe et al., 2001, Yaffe et al., 2009, Pignatti et al., 2002, Lee, S. et al., 2013, Rajan et al., 2015, van Gelder et al., 2004, Psaltopoulou et al., 2008, Jedrziewski et al., 2010). In brief, most studies had a measure of the volume of PA, Yaffe et al. (2001) addressed walking amount, Yaffe et al. (2009) addressed moderate or vigorous PA and Jedrziewski et al. (2010) studied the number of physical activities and the number of PA sessions. Of these prospective cohort studies launched in older age but with a lengthy follow-up over seven years, the most high-quality studies
(large cohort size at least of 500 participants, most important confounding factors taken into account, an accurate measure of PA) (van Gelder et al., 2004, Yaffe et al., 2001, Yaffe et al., 2009, Psaltopoulou et al., 2008, Lee, S. et al., 2013) provide support for the notion that PA protects from cognitive decline. The majority of these assessed the volume of PA, while Yaffe et al. (2009) specifically assessed the association of MVPA with later cognition. In these studies, light-intensity PA (Lee, S. et al., 2013), volume of PA (van Gelder et al., 2004, Yaffe et al., 2001, Psaltopoulou et al., 2008), and strenuous PA (Yaffe et al., 2009) were all associated with better cognitive function after a lengthy follow-up.


### 2.3.2 PHYSICAL ACTIVITY AND COGNITION: META-ANALYSES OF PROSPECTIVE COHORT STUDIES

Meta-analyses (Sofi et al., 2011, Morgan et al., 2012, Blondell et al., 2014, Guure et al., 2017) on PA and cognition have had similar estimates of association for high PA (see Table 1):

<table>
<thead>
<tr>
<th>Study</th>
<th>Risk ratio/OR/ effect size (95% CI)</th>
<th>Definition of PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sofi et al. 2011</td>
<td>0.62 (0.54–0.70)</td>
<td>High level of PA (the highest PA group regardless of the amount of PA categories</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in individual studies) compared to no PA or the lowest PA group</td>
</tr>
<tr>
<td>Morgan et al. 2012</td>
<td>0.66 (0.52–0.85)</td>
<td>PA compared to no PA or the highest PA level compared to “baseline”</td>
</tr>
<tr>
<td>Blondell et al. 2014</td>
<td>0.65 (0.55–0.76)</td>
<td>The highest level of PA compared to the lowest level of PA</td>
</tr>
<tr>
<td>Guure et al. 2017</td>
<td>0.67 (0.55–0.78)</td>
<td>High PA compared to low or no PA (if individual studies had more categories than</td>
</tr>
<tr>
<td></td>
<td></td>
<td>three, they were recategorized into three categories no/low, moderate, or high)</td>
</tr>
</tbody>
</table>

The methods to obtain the results did differ between the studies. Sofi et al. (2011), Morgan et al. (2012), and Blondell et al. (2014) used more traditional methods...
in their meta-analysis, while Guure et al. (2017) used Bayesian parametric and non-parametric methods. Sofi et al. (2011) and Morgan et al. (2012) did not assess the quality of the studies included systematically, but did perform sensitivity or subgroup analyses. Blondell et al. (2014) assessed the studies with a tool adapted from Singh (Singh et al., 2012), and Guure et al. (2017) assessed the study quality with the Meta-analysis of Statistics and Review Instrument (Joanna Briggs Institute, 2010). The number of cohorts included were 15, 9, 17, 22 for the meta-analyses from Sofi et al. (2011) Morgan et al. (2012), Blondell et al. (2014) and Guure et al. (2017), respectively. None of the meta-analyses found evidence of publication bias. Heterogeneity, on the other hand, was found by Morgan et al. (2012) and Blondell et al. (2014) but not by Sofi et al. (2011) and Guure et al. (2017). A common finding for all four of the meta-analyses was that shorter follow-ups were associated with greater positive associations, the finding being strongest in the meta-analysis from Guure et al. (2017): no significant association was found between PA and decreased cognitive decline incidence if only studies with follow-ups longer than 5 years were included. In the study from Sofi et al. (2011), studies with larger sample sizes and in the study from Blondell et al. (2014), high-quality studies and studies with at least 10 year follow-ups showed weaker or non-significant associations than smaller studies, studies with shorter follow-ups, or low-quality studies. In the meta-analysis from Guure et al. (2017), PA was positively associated with later cognition only in older cohorts at least 65 years old and follow-ups shorter than 5 years. These results from the subgroup analyses arouse the suspicion that the considerable associations found in the main results of these meta-analyses could be at least to some extent due to reverse causality. However, of the four very good quality prospective studies previously presented in this thesis with truly long-term follow-ups starting from midlife, all show a significant association between PA and cognition.

2.3.3 PHYSICAL ACTIVITY AND COGNITION: RANDOMIZED CONTROLLED TRIALS

Truly long-term RCTs with large sample sizes are difficult to carry out. Meta-analyses of RCTs assessing aerobic exercise and cognition in adults with healthy cognition demonstrate the difficulties in this field. The mean sample size in the trials of most of the meta-analyses in this field have ranged between 54 and 95 (Smith et al., 2010, Hindin et al., 2012, Kelly et al., 2014, Young et al., 2015, Northey et al., 2018, Barha et al., 2017). The length of the intervention has also been moderate in these meta-analyses, mostly less than one year (Smith et al., 2010, Hindin et al., 2012, Kelly et al., 2014, Young et al., 2015, Northey et al., 2018, Barha et al., 2017). In these meta-analyses of small studies with relatively short interventions, two reported beneficial effects on cognition (Smith et al., 2010, Hindin et al., 2012) and two did not (Young et al., 2015, Kelly et al., 2014). The most comprehensive
of these meta-analyses was that of Smith et al. (2010), covering 29 RCTs. The most recent meta-analysis in this field differs from these earlier meta-analyses in that it did not include studies with few participants but included only more robust larger trials (de Souto Barreto, Demougeot et al., 2018). The mean sample size in the trials of this meta-analysis was 520 and the mean duration 14.4 months. They concluded that evidence from RCTs is limited and does not support significant risk reduction for cognitive impairment with physical exercise. The single largest RCT with a lengthy duration (n=1476, duration: 2 years) with moderate-intensity PA showed no improvement in global or domain-specific cognitive function in elderly adults of 70–89 years of age (Sink et al., 2015).

The results from three robust multidomain RCTs (targeting not only exercise but also diet and/or cognitively stimulating activities) with large sample sizes (n=1190–3454) and with lengthy durations (2–6 years) have been published in recent years (Ngandu et al., 2015, Andrieu et al., 2017, Moll van Charante et al., 2016). PA was one of the intervention areas in FINGER (Ngandu et al., 2015) and in MAPT (Andrieu et al., 2017), whereas in PreDIVA (Moll van Charante et al., 2016) the specific lifestyle targets are not specified but most probably include PA as well. While the FINGER study (Ngandu et al., 2015) found a significant protective effect on cognition, MAPT (Andrieu et al., 2017) and PreDIVA (Moll van Charante et al., 2016) failed to do so. The results from FINGER have been criticized as being too weak (Lampit et al., 2015) and defended by the authors on the basis of being relevant in a public health context in the long-term (Kivipelto et al., 2015). On the other hand, the intervention in MAPT might have been too late for the study participants because 40% of the participants had very mild dementia according to the Clinical Dementia Rating at baseline (Kivipelto et al., 2017, Richard et al., 2012); the intervention did not result in an increase in PA levels (Andrieu et al., 2017). As for the PreDIVA study, it had considerably less rigorous interventions: only lifestyle advice, motivational interviewing techniques and five educational sessions, in contrast with several individual nutritional sessions, nutritional group sessions, 144 computer-training sessions and weekly guided exercise sessions in FINGER. The FINGER trial has also been criticized for exhibiting too large an effect due to a learning effect from computer-training affecting the cognitive outcomes (Kivimäki et al., 2015). The authors defended their study by describing how cognitive outcomes were measured with pen-and-paper assessment and with very different tests than the tasks used in computer-training (Kivipelto et al., 2015).

In a secondary analysis of the MAPT trial addressing the placebo group, LTPA was weakly associated with some of the cognitive measures used (de Souto Barreto, Andrieu et al., 2018). In the Finnish Prevention Study of Diabetes lasting approximately 4 years, cognition was followed up 9 years after the end of the intervention and the intervention did not result in cognitive benefits (Luchsinger
et al., 2015). In their secondary analysis, lifestyle factors were assessed individually and PA did not have a significant effect on cognition (Lehtisalo et al., 2016).

To conclude, evidence of a positive association between PA and cognition is scarce, controversial, and based on either small studies with relatively short durations or lengthier multidomain studies’ secondary analyses of PA. Convincing evidence to one direction or the other is still lacking.

Despite the insufficient evidence on the association of PA and cognition in cognitively healthy participants at baseline, promising evidence from cohorts with MCI (Heyn et al., 2004) or even dementia (Forbes et al., 2013, Groot et al., 2016) exist. According to a more recent review (Öhman et al., 2014), the evidence for dementia also has controversies and methodological shortcomings, but evidence for the positive association between PA and cognition was described as more rigorous.

2.3.4 OBJECTIVELY MEASURED PHYSICAL ACTIVITY AND COGNITION

To my knowledge, nine studies assessing the association between objectively measured PA and cognition have been published (see Table 2).
<table>
<thead>
<tr>
<th>Study</th>
<th>Mean age (if longitudinal age at baseline)</th>
<th>Cohort size</th>
<th>Longitudinal/ Cross-sectional (L/C)</th>
<th>PA measure</th>
<th>Outcome measure</th>
<th>Result</th>
<th>Sedentary behavior measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middleton 2011</td>
<td>74.8</td>
<td>197</td>
<td>L</td>
<td>Energy expenditure during 2 weeks using doubly labelled water</td>
<td>MMSE</td>
<td>Highest tertile of PA had lower odds of cognitive impairment than lowest tertile of PA</td>
<td>No</td>
</tr>
<tr>
<td>Buchman 2012</td>
<td>81.6</td>
<td>716</td>
<td>L</td>
<td>Wrist-worn accelerometer</td>
<td>A battery of 19 cognitive tests and clinician assessment if needed with NINCDS-ADRDA criteria</td>
<td>A higher level of total daily PA was associated with a reduced risk of AD</td>
<td>No</td>
</tr>
<tr>
<td>Zhu 2017</td>
<td>69.7</td>
<td>6452</td>
<td>L</td>
<td>Hip-worn accelerometer</td>
<td>A battery of cognitive tests (global cognition, executive function, memory)</td>
<td>MVPA associated with better memory, executive function, and global cognitive function</td>
<td>Yes, no association</td>
</tr>
<tr>
<td>Halloway 2017</td>
<td>66</td>
<td>174 at baseline, 55 at follow-up</td>
<td>L (change in PA with change in cognition)</td>
<td>Hip-worn accelerometer</td>
<td>A battery of cognitive tests assessing episodic memory, perceptual speed, and semantic memory</td>
<td>Maintaining MVPA (but not LPA) was associated with maintaining semantic memory, but not episodic memory or conceptual speed</td>
<td>No</td>
</tr>
<tr>
<td>Barnes 2008</td>
<td>83</td>
<td>2736 (only women)</td>
<td>C</td>
<td>Wrist-worn accelerometer</td>
<td>MMSE, Trail Making Test Part B*</td>
<td>Highest PA quartile had better cognitive function than lowest PA quartile</td>
<td>No</td>
</tr>
<tr>
<td>Wilbur 2012</td>
<td>66</td>
<td>174</td>
<td>C</td>
<td>Hip-worn accelerometer</td>
<td>A battery of tests of episodic memory and executive function (domains: inattention, interference control, word fluency)</td>
<td>MVPA correlated with interference control and word fluency but not with inattention or episodic memory</td>
<td>No</td>
</tr>
<tr>
<td>Kerr 2013</td>
<td>83</td>
<td>217</td>
<td>C</td>
<td>Hip-worn accelerometer</td>
<td>Trail Making Test Part A and B* (speed of processing, visual search speed, mental flexibility, executive function)</td>
<td>MVPA was associated with better processing speed and executive function, LPA not</td>
<td>No</td>
</tr>
<tr>
<td>Zhu 2015</td>
<td>70.1</td>
<td>7098</td>
<td>C</td>
<td>Hip-worn accelerometer</td>
<td>A battery of cognitive tests</td>
<td>MVPA was associated with better cognitive function, LPA not</td>
<td>Yes, no association</td>
</tr>
<tr>
<td>Johnson 2016</td>
<td>64.0</td>
<td>188</td>
<td>C</td>
<td>Hip-worn accelerometer</td>
<td>Trail Making Test Part A and B*</td>
<td>LPA was associated with better executive function, MVPA not</td>
<td>Yes, no association</td>
</tr>
</tbody>
</table>

The longitudinal studies all found significant association between either the amount of PA (Middleton, L. E. et al., 2011, Buchman et al., 2012) or MVPA (Zhu, W. et al., 2017) and cognition. The follow-ups were, nevertheless, short and cohorts in old age predispose these studies to be biased by reverse causation. Halloway et al. (2017) addressed the association of change in PA with change in cognition, but the follow-up rate in their study was 32% and maintaining MVPA (but not LPA) was associated with maintaining semantic memory but not episodic memory or perceptual speed.

The results of the cross-sectional studies were mostly in favor of a positive association between PA and cognition either for total PA (Barnes et al., 2008) or MVPA (Zhu, W. et al., 2015, Kerr et al., 2013). One study did not find an association between MVPA and cognition (Johnson, L. G. et al., 2016), but with LPA and cognition. However, in this study the cohort size was considerably smaller than in the other cross-sectional studies. One study assessed the correlation between LPA or MVPA with many subdomains of cognitive function (Wilbur et al., 2012). In their study, MVPA correlated with interference control and word fluency but not with episodic memory or inattention, while LPA correlated with inattention but not interference control, episodic memory, or word fluency. No study found an association between SB and cognition. That said, it is clear that evidence from studies assessing the association of PA and cognition with an objective measure of PA is still scarce.

To my knowledge, thus far no studies on objectively measured PA and cognition with genetically informative data have been published. Since both PA (Stubbe et al., 2006) and cognition (in adulthood) (Panizzon et al., 2014, Haworth et al., 2009) are, in moderate proportions, genetically determined, it is important to study the association and to take genetics into account.

2.3.5 PHYSICAL ACTIVITY AND DEMENTIA: PROSPECTIVE STUDIES

To the best of my knowledge, 11 observational studies addressing midlife PA and late-life dementia with a lengthy follow-up have been published (Carlson et al., 2008, Chang, M. et al., 2010, Elwood et al., 2013, Gelber et al., 2012, Gross et al., 2017, Morgan et al., 2012, Yamada et al., 2003, Rovio et al., 2007, Tolppanen et al., 2015, Rovio et al., 2005, Andel et al., 2008). One of these is a twin study with co-twin control design (n=147) (Carlson et al., 2008) and two of these are case-control studies (Andel et al., 2008, Gelber et al., 2012). The cohort sizes in the other studies ranged from 646 to 4945, the cohort size being smallest in the study by Gross et al. (2017) and largest in the study by Chang et al. (2010). Follow-up lengths were from 16 to 31 years. Only the study by Rovio et al. (2007) addressed only work-related PA, others studied LTPA or did not differentiate between work-related PA
and LTPA. All studies controlled for age and sex except for Gelber et al. (2012) and Elwood et al. (2013) who had only male participants. Education or a test of intelligence (Elwood et al., 2013) was controlled for in all of the studies except that of Gross et al. (2017) which was implemented in a cohort of medical graduates. At least some vascular risk factors were controlled for in all of the studies except for Yamada et al. (2003), Elwood et al. (2013) and Carlson et al. (2008). Yamada et al. (2003) and Carlson et al. (2008) used a crude measure of PA while other studies had more specific and thorough PA measures. The studies from Rovio et al. (2005) and Tolppanen et al. (2015) are from the same study population but with different categorizations, designs and emphases.

Studies addressing the volume of PA and dementia had different results: Morgan et al. (2012) found no association, Gelber et al. (2012) and Andel et al. (2008) found a significant inverse association, and Chang et al. (2010) found a significant association only for moderate PA, not high but on the other hand, the direction of the association was similar for high PA and there were very few cases of dementia in the high PA group, affecting the significance levels. Gross et al. (2017), Rovio et al. (2005), and Tolppanen et al. (2015) did not address the light end of the spectrum of PA, but studied PA causing perspiration or MET-hours of moderate and vigorous intensity. Rovio et al. (2005) found a significant association between PA and decreased incidence of dementia, Gross et al. (2017) did not while Tolppanen et al. (2015) found a significant inverse association for moderate PA only in men or only if late-life self-reports of PA were included, creating a risk of reverse causation. Carlson et al. (2008), Chang et al. (2010) and Yamada et al. (2003) had PA indices constructed from PA of both LPA (e.g., gardening) and vigorous activity and did not differentiate between these two. Carlson et al. (2008) and Yamada et al. (2003) found no significant association, while Chang et al. (2010) did find a significant inverse association. Rovio et al. (2007) reported no significant association between work-related PA and dementia.

In conclusion, the most high-quality studies (baseline in midlife, a follow-up of over 10 years, no significant selection bias in the cohort, a thorough measure of PA, the result controlled for at least age, sex, education or a measure of intelligence and vascular morbidity) of LTPA or PA in general show inconsistent results (Chang, M. et al., 2010, Gelber et al., 2012, Morgan et al., 2012, Tolppanen et al., 2015, Rovio et al., 2005, Andel et al., 2008). The studies from Gelber et al. (2012), Andel et al. (2008) and Rovio et al. (2005) show significant inverse associations between LTPA and dementia, the study from Morgan et al. (2012) does not, while the studies from Chang et al. (2010) and Tolppanen et al. (2015) are controversial in showing a significant inverse association only for moderate amounts but not for high amounts (Chang, M. et al., 2010) or only for men (Tolppanen et al., 2015). In addition to these studies assessing midlife PA and dementia, many studies have assessed the association of late-life PA and dementia (Appendix II). These studies may be subject
to reverse causation since dementia is a slowly developing degenerative disease and in late life the preclinical process may already affect PA participation.

2.3.6 PHYSICAL ACTIVITY AND DEMENTIA: META-ANALYSES OF PROSPECTIVE STUDIES

Earlier meta-analyses of PA and all-cause dementia have shown quite similar pooled RRs for high PA (see Table 3) (Hamer et al., 2009, Morgan et al., 2012, Blondell et al., 2014, Guure et al., 2017, Xu et al., 2017). Apart from the study from Hamer & Chida (2009) who did not perform a sensitivity analyses, all these earlier meta-analyses report either significant risk of publication bias or that there was no significant association if: 1) follow-up is longer than 10 years or 2) the baseline age of the participants is under 65 years. Even if the conclusions in all these earlier meta-analyses except Morgan et al. (2012) support the hypothesis that PA protects from dementia, the results from the subgroup analyses arouse the suspicion that the associations found in the main results of these meta-analyses could, at least to some extent, be due to reverse causality. In the study from Xu et al. (2017), the inverse association of PA and dementia was, however, also nearly statistically significant in the subgroup of younger participants (< 65 years) and also significant in longer (> 10 years) follow-ups. The most recent meta-analysis assessing physical inactivity and dementia supports this notion that there is no convincing evidence for an association between physical activity and dementia when only follow-ups longer than 10 years are included (Kivimäki et al., 2019). Their results were adjusted for most important confounding factors (age, sex, education or socio-economic status and vascular morbidity). Although they had 19 study results included in their meta-analysis, this information came from only 9 different prospective cohort studies at different time points.
The meta-analyses addressing PA and AD all point to an inverse association (Table 4) (Beckett et al., 2015, Beydoun et al., 2014, Daviğlus et al., 2011, Santos-Lozano et al., 2016, Guure et al., 2017), and some postulate that the association is especially strong for AD compared to all-cause dementia or vascular dementia (Beydoun et al., 2014, Guure et al., 2017). The meta-analyses that incorporated a quality assessment in their study found good quality in all studies included (Santos-Lozano et al., 2016, Guure et al., 2017).

<table>
<thead>
<tr>
<th>Meta-analysis</th>
<th>Pooled RR (95% CI)</th>
<th>Number of studies included</th>
<th>Heterogeneity</th>
<th>Evidence for publication bias</th>
<th>Subgroup analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamer et al. 2009</td>
<td>0.72 (0.60–0.86)</td>
<td>16</td>
<td>Yes</td>
<td>No</td>
<td>No dose-response</td>
</tr>
<tr>
<td>Morgan et al. 2012</td>
<td>0.78 (0.65–0.94)</td>
<td>16</td>
<td>Yes</td>
<td>Yes</td>
<td>Similar associations for studies with short and long follow-ups</td>
</tr>
<tr>
<td>Blondell et al. 2014</td>
<td>0.86 (0.76–0.97)</td>
<td>21</td>
<td>Yes</td>
<td>Maybe</td>
<td>Significant inverse association but only in studies with follow-ups shorter than 10 years</td>
</tr>
<tr>
<td>Xu et al. 2017</td>
<td>0.73 (0.62–0.87)</td>
<td>15</td>
<td>Yes</td>
<td>No</td>
<td>Dose-response present, but association was significant only in studies with elderly adults (≥65 years)</td>
</tr>
<tr>
<td>Guure et al. 2017</td>
<td>0.79 (0.69–0.88)</td>
<td>32</td>
<td>No</td>
<td>No</td>
<td>Association significant only in studies with elderly adults (≥65 years)</td>
</tr>
<tr>
<td>Kivimäki et al. 2019</td>
<td>1.01 (0.89–1.13)</td>
<td>19</td>
<td>No</td>
<td>Not evaluated</td>
<td>Association significant only in studies &lt; 10 years of duration</td>
</tr>
</tbody>
</table>
Table 4  Meta-analyses of prospective studies assessing physical activity and AD

<table>
<thead>
<tr>
<th>Meta-analysis</th>
<th>Pooled RR (95% CI)</th>
<th>Number of studies included</th>
<th>Heterogeneity</th>
<th>Evidence for publication bias</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davigli et al. 2011</td>
<td>0.72 (0.53–0.98)</td>
<td>9</td>
<td>Yes</td>
<td>Not told</td>
<td>Insufficient evidence because of self-reporting PA and heterogeneity in the results</td>
</tr>
<tr>
<td>Beydoun et al. 2014</td>
<td>0.58 (0.49–0.70)</td>
<td>8</td>
<td>Yes</td>
<td>No</td>
<td>PA protects from AD</td>
</tr>
<tr>
<td>Beckett et al. 2015</td>
<td>0.61 (0.52–0.73)</td>
<td>9</td>
<td>No</td>
<td>Not told</td>
<td>PA seems to decrease AD incidence, but reverse causation is possible due to short follow-ups</td>
</tr>
<tr>
<td>Santos-Lozano et al. 2016</td>
<td>0.65 (0.56–0.74)</td>
<td>10</td>
<td>Moderate</td>
<td>No</td>
<td>PA might prevent AD, caution must be warranted due to the heterogeneity and non-specificity of PA measurement methods although quality of the studies included was high according to their quality assessment</td>
</tr>
<tr>
<td>Guure et al. 2017</td>
<td>0.62 (0.49–0.75)</td>
<td>17</td>
<td>Yes</td>
<td>No</td>
<td>PA protects from AD and the association seems to be especially strong compared to all-cause dementia, vascular dementia, and cognitive decline, even in long follow-ups (&gt; 5 years)</td>
</tr>
</tbody>
</table>

Evidence on PA and vascular dementia does not suggest any significant association (Aarsland et al., 2010, Guure et al., 2017). Aarsland et al. (2010) found a significant association but the meta-analysis was limited by insufficient control of confounding factors. On the other hand, Guure et al. (2017) found no significant association for PA and VaD, and according to the Meta-Analyses of Statistics Assessment and Review Instrument (Joanna Briggs Institute) the methodology in the eight studies from Guure et al.’s (2017) meta-analysis was high. However, it must not be forgotten that studies addressing this issue are so far fairly scarce.

In addition to the meta-analyses described, a plethora of narrative and systematic reviews are conducted addressing PA and dementia. Because many are outdated or do not use systematic methods, they are not discussed. One systematic review on PA and AD, however, was recently published and merits a closer look. The authors of this study used an adapted quality assessment tool specifically targeting prospective studies of PA and AD (Stephen et al., 2017). Instead of the vague quality criteria often used, they had very specific, detailed and logical quality criteria. Of the 24 prospective observational studies assessing PA and AD, 16 were of moderate quality and 8 of low quality. According to their robust systematic review, they conclude an inverse association between PA and AD with moderate quality of evidence. They point out that many of the studies failing to find this association did not separate
between LTPA and work-related PA, which might explain their different results. They also go on to suggest that the window of opportunity is broad regarding age.

According to up-to-date evidence, the inverse association between PA and all-cause dementia or AD seems clear but is less so for PA and VaD. The possibility of reverse causality, however, still remains because the evidence comes predominantly from studies with fairly short follow-ups and with baseline in old age. Nevertheless, in the study from Xu et al. (2017), the association was also significant in the follow-ups longer than 10 years and also near the significance level in the subgroup of participants younger than 65 years. The contribution of genes in the association is unclear and the intensity and volume of PA needed to produce protective benefits also remains ambiguous.

2.3.7 CARDIORESPIRATORY FITNESS AND COGNITION

As discussed previously in this thesis, PA has been associated with better cognition later in life (Chang, M. et al., 2010, Elwood et al., 2013, Singh-Manoux et al., 2005, Virta et al., 2013a). However, people who exercise more are often more fit and since we know cardiorespiratory fitness is greatly determined by genetics (Fagard et al., 1991, Sundet et al., 1994), one can wonder whether it is really PA that affects cognitive aging or cardiorespiratory fitness and the underlying genetics?

The association of cardiorespiratory fitness and cognition is a much less studied topic than that of PA and cognition. The only meta-analysis specifically addressing the association of cardiorespiratory fitness and cognition is outdated and the studies in it have very small sample sizes, mostly under 30 participants (Etnier et al., 2006). This meta-analysis does suggest a significant association between cardiorespiratory fitness and cognition, but the results must be interpreted with caution because of the small sample sizes in the studies included. More recent cross-sectional studies, also with mainly quite small sample sizes, also suggest an association between cardiorespiratory fitness and cognition (Brown, A. D. et al., 2010, Etnier et al., 2007, Netz et al., 2011, Newson et al., 2006, Scott et al., 2016), as do larger longitudinal studies (Barnes et al., 2003, Defina et al., 2013, Kulmala et al., 2014, Liu et al., 2012, Wendell et al., 2014). Controversies about the cognitive domain affected and about the relationship to exercise exist. A meta-analysis of fitness-enhancing RCTs suggests that fitness especially improves executive control (Colcombe, S. et al., 2003), while a longitudinal study suggests that the domain most affected by fitness is memory (Wendell et al., 2014). A recent RCT suggests that cardiorespiratory fitness change predicts cognition changes better than simple participation in PA (Vidoni et al., 2015), but a meta-analysis of many RCTs shows no cognitive benefits from exercise even if a concurrent cardiorespiratory fitness change is shown to occur (Young et al., 2015).
2.3.8 RESISTANCE TRAINING AND COGNITION

It has been suggested that resistance training may enhance cognitive performance via different mechanisms than aerobic exercise (Liu-Ambrose et al., 2009). Evidence on the association of resistance training and cognition is, however, scarce. No prospective cohort studies have addressed the issue to my knowledge, but several RCTs on resistance training and cognition have been published. Many of the reviews and meta-analyses of RCTs point out that there is a lack of high-quality studies (van Uffelen et al., 2008, Chang, Y. K. et al., 2012, Saez de Asteasu et al., 2017). Many of them suggest that resistance training might have a beneficial cognitive effect on at least some domains of cognition, perhaps even on many (Chang, Y. K. et al., 2012, Liu-Ambrose et al., 2009, Saez de Asteasu et al., 2017, Northey et al., 2017, Kelly et al., 2014). The cognitive domain which seems to be most affected by strength training is probably executive control (Northy et al., 2017, Saez de Asteasu et al., 2017), but also possibly memory and working memory (Northy et al., 2017). Additionally, contradicting conclusions have been made (van Uffelen et al., 2008, Snowden et al., 2011). Sáez et al. (2017) suggest that strength training must be at least moderate-to-high intensity and progressive to provide cognitive benefits. They also discovered that attendance rate was a significant moderator in their meta-analysis and low attendance rate at least partly explained the null results (Saez de Asteasu et al., 2017). It must be noted that the assumptions made are based on only a few studies with altogether approximately 304–409 participants in the meta-analyses or pooled resistance training groups (Kelly et al., 2014, Saez de Asteasu et al., 2017), and sample sizes in individual studies are small, on average 74 (Northy et al., 2017).

The suggested pathways through which resistance training may enhance cognitive performance are insulin-like growth factor (IGF-1) and through homocysteine concentration. Indeed, the cognitive benefits from aerobic exercise have been associated with elevating the levels of BDNF and strength training has been shown not to elevate the levels of BDNF (Huang, T. et al., 2014). Contradicting evidence, however, exists. Marston et al. (2017) showed that intense resistance training can increase the peripheral levels of brain-derived neurotrophic factor (Marston et al., 2017).

In conclusion, there are few studies on the topic and even fewer high-quality studies, the procedures and exercise protocols are very heterogenous and decisive conclusions cannot be made. In the first large-scale longer-term multidomain intervention against cognitive decline, resistance training was one of the targets of the intervention and the intervention showed significant cognitive improvements (Ngandu et al., 2015). Another large multidomain intervention did provide individual home-based training programs for the participants, but the PA levels of the participants did not increase after baseline and the intervention as a whole did not have significant cognitive benefits (Andrieu et al., 2017).
2.3.9 MECHANISTIC STUDIES

Both animal models and human studies have revealed several modes of action regarding how aerobic exercise may work to improve cognition and buffer against dementia. Aerobic exercise has been shown to increase neuroplasticity (Ding, Vaynman, Souda et al., 2006, Huttenrauch et al., 2016), induce macro-scale changes such as larger volume of the frontal lobe (Colcombe, S. J. et al., 2006) and hippocampus (Erickson et al., 2011), prevent accumulation of amyloid plaques (Adlard et al., 2005, Herring et al., 2016), and reduce oxidative stress (Radak et al., 2016) and neuroinflammation (Nichol et al., 2008). Additionally, aerobic exercise can work through other indirect ways for the benefit of cognition, such as through cardiovascular factors and through reducing stress (Atlantis et al., 2004) or enhancing the quality and lengthening the duration of sleep (Dolezal et al., 2017).

In molecular language, learning and memory mean efficient communication between neural cells and enhancement of appropriate synaptic connections. Exercise has been shown to facilitate long-term potentiation (i.e., increased neuron response which is crucial to learning) (Patten et al., 2013), up-regulate proteins involved in neural plasticity (Ding, Vaynman, Souda et al., 2006), up-regulate genes associated with synaptic plasticity (Huttenrauch et al., 2016), increase dendrite length, dendrite complexity, and dendrite spine density in the dentate gyrus of the hippocampus (Eadie et al., 2005), and to up-regulate genes coding N-methyl-D-aspartate receptors, which are important for synapse strengthening and learning (Molteni et al., 2002).

One important mediator between exercise and the enhancement in LTP and the strengthening of synapses is BDNF (Vaynman et al., 2006). PA was first found to increase the levels of BDNF in rats’ brains in 1996 (Neuper et al., 1996), and only much later to be associated with peripheral BDNF levels in humans (Huang, T. et al., 2014). Some obscurities, however, are associated with the issue. Although, the peripheral levels of BDNF have been elevated in experimental studies following acute or chronic exercise, a contrary relationship has been found in observational studies: habitual PA along with cardiorespiratory fitness was associated with decreased levels of peripheral BDNF (Huang, T. et al., 2014). Despite the plethora of studies showing an association between exercise and higher peripheral BDNF levels (Huang, T. et al., 2014, Coelho et al., 2013, Griffin et al., 2011) and more studies showing an association between low peripheral BDNF levels and lower cognition (Leckie et al., 2014) or dementia (Weinstein et al., 2014) and even depression (Molendijk et al., 2014), schizophrenia (Fernandes, B. S. et al., 2015) and autism (Hashimoto et al., 2006), it is still unclear how well the peripheral levels of BDNF actually reflect the brain levels of BDNF. Two hypotheses have been proposed regarding where serum BDNF originates from: mainly from the brain (Krabbe et al., 2007, Rasmussen et al., 2009) or mainly from platelets and megakaryocytes (Fujimura et al., 2002).
According to a recent synopsis on the peripheral levels of BDNF from a Finnish study group studying depression, the changes in serum BDNF first and foremost reflect altered release from the platelets (Castren et al., 2017). However, they speculate that because there are similarities in the BDNF synthesis in megakaryocytes and neurons, there might also be parallels between brain and serum BDNF content and release. They also point out the difficulties in studying this field: the human brain levels of BDNF are out of reach while, for example, mice express barely detectable levels of BDNF in megakaryocytes making it impossible to compare the mechanisms and sources of blood BDNF in humans and in mice.

In animal models, PA has also been associated with increases in other neuronal growth factors: nerve growth factor (Ang et al., 2003, Neep et al., 1996), insulin-like growth factor I (IGF-1) (Ding, Vaynman, Akhavan et al., 2006, Llorens-Martin et al., 2010), and vascular endothelial derived growth factor (Schobersberger et al., 2000, Llorens-Martin et al., 2008). These neurotrophins (BDNF, IGF-1, VEGF) have been thought to be essentially implicated in synaptogenesis and angiogenesis in the brain.

Aerobic exercise has been shown to increase vasculature in the rat brain, particularly in the hippocampus (Lopez-Lopez et al., 2004). The longevity and the regional selection of the changes remain uncertain (Van der Borght et al., 2009), as well the exact age-dependency of the effect (van Praag et al., 2005). Angiogenesis and increased vasculature in the brain facilitates and enables a more efficient oxygen, glucose and neurotrophic factor supply to the brain. In humans, the effect of exercise on cerebral blood flow has been examined with $^{133}$Xe-tracers (Rogers et al., 1990), ultrasound (Ainslie et al., 2008), PET and MRI measuring either cerebral blood flow (Pereira et al., 2007, Chapman et al., 2013, Tarumi et al., 2013, Maass et al., 2015, Brown, A. D. et al., 2010, Hiura et al., 2014, Hiura et al., 2018b, Hiura et al., 2018a) or vasculature with MR angiography (Bullitt et al., 2009). With these techniques, PA has been shown to be associated with better cerebral perfusion in humans (Rogers et al., 1990, Ainslie et al., 2008, Tarumi et al., 2013, Pereira et al., 2007, Chapman et al., 2013) age-dependently (Maass et al., 2015) and with an increase in small-caliber vessels and lower vessel tortuosity in the human brain (Bullitt et al., 2009). However, the acute effect of exercise on regional cerebral blood flow seems to be dependent on the exercise phase, according to recent studies deploying PET imaging (Hiura et al., 2014, Hiura et al., 2018b, Hiura et al., 2018a).

In rats, exercise has even been shown to increase the rate of neurogenesis in the hippocampus (van Praag et al., 1999), provided that it is aerobic and sustained (Nokia et al., 2016). In humans, it is unclear whether adult neurogenesis is possible. There have been claims for (Spalding et al., 2013) and against it (Sorrells et al., 2018).

PA has been associated not only in microscale structural changes in the brain, but also in macroscale structure changes of the brain. Studies imply that PA increases grey matter volume (Erickson et al., 2010, Rovio et al., 2010) and more specifically
the volume in the frontal lobes (Colcombe, S. J. et al., 2006) and hippocampus (Thomas et al., 2016, Erickson et al., 2011). According to a recent review, 82% of brain grey matter areas are modifiable by PA (Batouli et al., 2017). Although being a very inspiring finding, a few criticisms must be pointed out: no quality assessment was included and combining all the positive associations found in various studies and not reporting the negative findings or null findings can and most probably will lead to an overly optimistic conclusion. In young adult healthy monozygotic twins with similar backgrounds but discordant LTPA for the preceding three years, there were no significant differences between total grey matter volume and total white matter volume, but regionally the striatum (an area involved in motor control and movement) and prefrontal cortex (involved in executive function) were significantly larger in physically active twins (Rottensteiner et al., 2015). These findings indicate that the PA’s influence in the brain is site- and task-specific, instead of being a general cognitive booster.

As for white matter in the brain, both cross-sectional and intervention studies have been published implying that PA is associated with better integrity (Gons et al., 2013, Voss et al., 2013) and fewer white matter lesions (Fleischman et al., 2015), but conflicting evidence also exists (Rovio et al., 2010). A recent review concludes that the area is still in its infancy and that although there seems to be a positive correlation between PA and increased white matter volume, better white matter microstructure, and fewer white matter lesions, caution is warranted due to small effect sizes and the number of negative findings (Sexton et al., 2016).

PA may also beneficially influence cognition via affecting cardiovascular risk factors such as obesity, hypertension, diabetes, and dyslipidemia. Better sleep (Dolezal et al., 2017), reduced stress (Atlantis et al., 2004), the joys of social interaction (Vancampfort et al., 2017), augmented mood (Peluso et al., 2005), and decreased incidence of dementia (Waller et al., 2016) may also work as vectors conveying at least in part the beneficial effects of PA.

Low-grade inflammation measured with high-sensitivity CRP (C-reactive protein) has been linked to dementia: both VaD and AD (Schmidt et al., 2002). PA has been shown to reduce the peripheral levels of proinflammatory cytokines in humans (Nascimento et al., 2014) and to reduce the levels of proinflammatory cytokines IL-1β and tumor necrosis factor alpha (TNF-α) specifically in the (mouse) hippocampus (Nichol et al., 2008). Free radicals can cause oxidative damage in the cell by affecting proteins, receptors, and even factors regulating gene expression. Regular moderate exercise induces adaptive changes to buffer against oxidative stress both in the blood (Robertson et al., 1991) and in the (rat) brain (Radak et al., 2016). PA may also influence the gut microbiome of the host (Codella et al., 2017) and since gut microbiome is associated with dementia among other things (Vogt et al., 2017), this is one possible mechanism for how PA might provide protection against dementia.
In animal models, exercise directly reduces cerebral Aβ levels and amyloid deposition (Adlard et al., 2005, Ambree et al., 2006, Lazarov et al., 2005), even in advanced stages of AD (Herring et al., 2016). While exercise in young rodents targets primarily amyloidogenic APP processing and Aβ degradation, long-term running in older mice with advanced stages of AD affects Aβ clearance across the blood–brain barrier, autophagy machinery and anti-inflammatory processes coupled with antioxidative effects (Herring et al., 2016). The reduction in cerebral amyloid plaque formation induced by exercise seems dose-dependent in mice (Moore et al., 2016, Herring et al., 2016). In humans, positron-emission imaging and the measurement of Aβ\textsuperscript{42} and tau from cerebral spinal fluid has enabled the in vivo examination of cerebral amyloid accumulation or estimation of the amount of AD pathology. The inspiring results from animal studies that PA reduces cerebral amyloid deposition have not been replicated directly in humans. In humans, PA has been associated with lower amyloid burden in two studies (Liang et al., 2010, Müller et al., 2018), has not been associated with amyloid burden in two studies (Vemuri et al., 2016, Souto Barreto et al., 2015), has been associated with decreased amyloid burden only in the subgroup of APOE e4 carriers but not non-carriers (Brown, B. M. et al., 2013, Head et al., 2012) and has been associated with decreased amyloid burden only in the subgroup of those already accumulating AD pathology among carriers of the autosomal dominant AD mutation (Brown, B. M. et al., 2017). In other words, conflict between results prevails.

While genes in the deoxyribonucleic acid (DNA) code for the phenotypic traits in living creatures, the environment can affect or modulate the gene expression with epigenetic changes. Epigenetic changes are modifications in the chromatin structure of DNA, meaning that DNA is wrapped up differently, affecting DNA expression. Accumulating evidence shows that exercise has a remarkable capacity to alter and regulate synaptic and cognitive plasticity through such epigenetic mechanisms as histone modification or DNA methylation (Fernandes, J. et al., 2017).

It has been suggested that exercise also builds up cognitive reserve. Indeed, many studies show that PA in young age is associated with better academic performance (Singh et al., 2012, Haapala et al., 2017a, Haapala et al., 2018) and with better executive function (Verburgh et al., 2014). Building up a high-functioning, efficient and resilient brain at young age could, thus, buffer against neuropathology in old age. The field is, however, still controversial. A recent systematic review and meta-analysis concluded that the evidence for a positive association between childhood PA and better academic performance and cognitive function is limited and equivocal (Li, J. W. et al., 2017), and the direction of the causality could also be the opposite (Aaltonen et al., 2016).
2.4 TWIN STUDIES

2.4.1 DNA AND GENETIC SELECTION

In the cell nucleus, the code for living is preserved in a structure called chromatin, encompassing the DNA sequence (the actual code) and proteins compressing DNA in a tight space and regulating its expression. Only four bases in DNA code the vast information for different human traits and the function of proteins and cells. Additionally, genetically inherited diseases or risk genes with low penetrance are encoded in the DNA. The genome, thus, carries a huge amount of information. When doing research on various environmental factors, the effect of genetic selection has to be considered. In successful double-blinded randomized trials, little genetic selection can be presumed. The problem in randomized trials is, however, that their duration is quite limited and long-lasting environmental exposures or lifestyle choices are difficult to study. In cohort studies, the effect of genetic or other selection is always an issue. If a cohort is cut in half according to one trait or lifestyle choice and its effect on the outcome is studied, it is always crucial to examine what other traits or lifestyle choices are present in the other half. Even if other lifestyle choices are carefully taken into account, genetic selection can still be hidden in the association.

2.4.2 CLASSICAL TWIN DESIGN

A twin study is a unique opportunity to study cause-and-effect relationships by taking the genome and early environmental exposures into account. Monozygotic twins (MZ) share their full DNA sequence including all segregating genes, while dizygotic twins (DZ) share on average half of their segregating genes. Variances in MZ and DZ twins enable the study of environmental and genetic contributions to individual differences of different traits and diseases (Polderman et al., 2015). If MZ twins are more similar than DZ twins with respect to some phenotypic trait, it is an indication that genetic influences affect the phenotype of that trait (Neale et al., 1992). The classical twin study compares the phenotypic resemblance of MZ and DZ twins (van Dongen et al., 2012). The sources of variation in individual traits are contributed by additive genetic influences, non-additive genetic influences, shared environmental factors, and unique environmental factors (Rijsdijk et al., 2002). The correlations between MZ and DZ twins are used to quantify the heritability of a trait and the effect the shared environmental factors have on a trait. This area of research is called quantitative genetics.
2.4.3 CO-TWIN CONTROL STUDIES

An ideal setting for an intervention study is a case-control study with MZ twins (Gesell, 1942). MZ twins’ genomes are identical and their childhood environment is nearly exactly the same, thus, these confounding factors are automatically controlled for (Boomsma, D. et al., 2002). Another type of co-twin study is observational co-twin studies in which MZ twins sharing all of their genes and very similar environmental backgrounds are studied observationally regarding one specific discordant behavior, for example LTPA (Kujala et al., 2002). These kinds of studies are valuable because they provide valuable information that controls for genetic factors.

2.4.4 THE ASSUMPTIONS OF TWIN STUDIES

When performing quantitative genetics, certain assumptions are made (see Table 5) (Silventoinen et al., 2008). These same assumptions are also relevant for co-twin control studies, but since co-twin control studies aim to distinguish solely the rough existence of genetic influences in the background of a trait, small violations of these assumptions hardly undermine the basic principle of these studies. In quantitative genetics, specific numeric contributions of genetic effects, shared environmental effects and unique environmental effects are pursued and, thus, violations of these assumptions are more influential.

Table 5 The assumptions of twin studies*

1. Absence of assortative mating

2. Absence or minimal effect of gene–environment interaction meaning that exposure to different environments is not random, but an individual with a certain genotype is more prone to end up in certain environments or experience certain kinds of life experiences. It can also be passive via their parents’ biology offering certain kinds of environments with greater probability. It can also mean that individuals with certain genotypes are more vulnerable to certain environmental effects than others (Figure 6).

3. Generalizability of twins to the general population. This denotes that twins are not significantly different from singletons for the trait being studied.

4. Equal environments assumption assumes that MZ and DZ twins and co-twins are equally correlated in their exposure to environmental factors.

* The information in the box is based on the article by Silventoinen et al. (2008)
Assortative mating has been proven to occur for traits such as height, BMI (Silventoinen et al., 2003) and intelligence (Escorial et al., 2012). However, Aarnio et al. (1997) showed in the Finnish population that regarding LTPA, the assumption of random mating holds (Aarnio et al., 1997). Gene–environment interaction is shown to occur, for example, in Dutch twins: upbringing reduced genetic influences on disinhibition (Boomsma, D. I. et al., 1999). The equal environments assumption has been criticized (Guo, 2001) due to parents treating MZ twins more similarly than DZ twins, but it can contradicted by claiming that this is due to the similarity of MZ twins (Silventoinen et al., 2008). The generalizability assumption is quite well achieved because despite being born prematurely more often and weighing less than singletons (Martin et al., 2003, MacGillivray et al., 1988), twins are proven to have similar mortality as the general population (Christensen et al., 1995, Kaprio, 2013) and the same prevalence of most diseases as singletons (Kyvik, 2000), and they are shown not to differ from singletons regarding personality traits of behavior (Johnson, W. et al., 2002, Pulkkinen et al., 2003). However, low birth weight may convey an increased risk for cognitive decline (Mosing et al., 2018). In fact, intrauterine conditions like maternal nutrition, stress, parity, and size affect the offspring’s phenotype and epigenetic state; the metabolic path to future outcomes may already be programmed, to some extent, during the fetal period (Gluckman et al., 2011) or early developmental stages (Lahiri et al., 2009).

In conclusion, assortative mating, a possible increased risk for cognitive decline due to lower birth weight and the possibility of gene–environment interaction have to be taken into account in the interpretation of twin studies on PA and cognition.
and dementia. But again, the effect of lower birthweight and small violations in the assumptions of absent assortative mating and absent or minimal effect of gene–environment interactions is small in the co-twin control design and more influential in quantitative genetics.

2.5 SUMMARY OF THE LITERATURE REVIEW

Numerous studies on the association of PA and cognition or dementia have been published. The majority of them have found a significant negative association between PA and cognition or dementia. Most of them have also been of short duration with only a few years of follow-ups and implemented in aged cohorts. Because the preclinical phase of dementia is most probably decades (Bateman et al., 2012, Fagan et al., 2014), short follow-ups and implementation in aged cohorts makes these studies liable for reverse causation. There is a lack of high-quality studies starting in midlife with over 10 years follow-up. The high-quality studies of the association of PA and cognition have uniformly shown a significant positive association between PA and cognition (Chang, M. et al., 2010, Elwood et al., 2013, Singh-Manoux et al., 2005, Virta et al., 2013a), but whether the association is the effect of lower BMI instead of PA is unclear based on these studies (only Chang et al. (2010) adjusted their results for BMI). The high-quality studies of PA and dementia are more controversial in showing a significant inverse association (Andel et al., 2008, Gelber et al., 2012, Rovio et al., 2005), no significant association (Morgan et al., 2012) or a significant association only in men (Tolppanen et al., 2015) or only for moderate but not high amounts of PA (Chang, M. et al., 2010). All of these prospective studies except for Gelber et al. (2012) adjusted their results for BMI, hence it seems likely that BMI does not mediate the findings. Robust RTCs with considerable sample sizes and long interventions are scarce (Smith et al., 2010), and the largest RCT thus far, to my knowledge, did not show any significant difference in cognitive function between the PA group and control group that received health education (Sink et al., 2015). A recent thesis study showed that cognition and mobility are closely related across the life course, and in old age, executive function predicts mobility better than mobility predicts executive function in 2-year follow-up (Poranen-Clark, 2018). This thesis study highlights the important issue: In young age, adulthood and old age, is PA the consequence of higher general cognitive abilities or does PA independently preserve cognition and prevent dementia? Or is genetic pleiotropy (i.e., one gene affecting two or more seemingly unrelated phenotypic traits) an important common denominator behind the association of PA and cognition?

Indeed, because the heritabilities for LTPA (48–71%) (Stubbe et al., 2006), all-cause dementia (43%) (Gatz et al., 1997) and general cognitive abilities (50%)
(Haworth et al., 2009) have been estimated to be considerable, it is important to study whether genetics account to some or a great extent for the possible association between PA and cognition or dementia. So far, few twin studies have been published and in these twin studies no significant associations between PA and cognition or dementia were found in the pairwise analyses comparing the twins against their co-twins (Andel et al., 2008, Carlson et al., 2008, Virta et al., 2013a).

The majority of the studies assessing the association of PA and cognition or dementia have used questionnaires or interviews to assess PA levels. There is also a lack of studies assessing the relationship of PA and cognition with an objective measure of PA. In addition, studies assessing the association of SB and cognition are a rarity.
3 THE AIMS OF THE STUDY

The objective of this thesis was to investigate whether PA is associated with decreased dementia mortality and decreased incidence of cognitive decline, and whether the possible association is explained by genetic factors. Additionally, this thesis aimed to study the relationship between objectively measured late-life PA and late-cognition and whether the possible association is also detected when the results control for genetics and childhood shared environment. The specific study aims were to assess whether:

1. midlife LTPA is associated with late-life dementia mortality. (Study I)
2. midlife LTPA is associated with late-life cognition. (Study II)
3. objectively measured late-life PA and sedentary time is associated with late-life cognition. (Study III)
4. genetic selection and childhood shared environment explain the possible associations observed. (Study I, II, III)
4 PARTICIPANTS AND METHODS

4.1 STUDY DESIGN AND PARTICIPANTS

The Finnish Twin Cohort was established in 1975 (Kaprio et al., 1978b). It is a longitudinal twin study and comprises all same-sex twins born before 1958 and with both co-twins alive in 1967 (Kaprio et al., 2002). The twins were selected from the Central Population Registry of Finland in 1974 and zygosity was determined by a validated questionnaire method inquiring about the similarity of childhood appearances and about strangers confusing the twins in childhood (Sarna et al., 1978). Later on, DNA samples have been used to verify zygosity and in Study II, the zygosity has been confirmed to be the same as with the questionnaire method in 98.4% and in Study III, for 99.3% of all twin individuals (when MZ and DZ zygosity has been taken into account, but not changes from the original group of unknown zygosity to MZ or DZ zygosity).

Four waves of data collection have been conducted. The surveys have included a comprehensive series of questions relating to health, lifestyle, and occupation and were administered in 1975, 1981, 1990, and 2011. The response rates for the surveys have been 89% in 1975, 84% in 1981, 77% in 1990, and 72% in 2011 (Kaprio et al., 1988, Piirtola et al., 2017, Kaprio, 2013). The rate of DZ twinning was high in Finland before the 1950s and is reflected in the Finnish Twin Cohort in the two-fold ratio of DZ: MZ twins (Rose et al., 1988). The Finnish Twin Cohort Study can be considered a good representation of the Finnish population because it comprises a very large population sample of Finnish twins, the response rates have been high in the conducted surveys, and, furthermore, mortality (Kaprio, 2013), cancer (Verkasalo et al., 1999), and the incidence rate of type 1 diabetes (Hyttinen et al., 2003) have been shown to be consistent with those in the general Finnish population. In 1996, the Finnish Twin Cohort Study was expanded to also include opposite-sex twins born in 1938–1949 (Kaprio et al., 2002).

In Study I, all twin individuals who had answered questionnaires from 1975 and 1981 adequately to be able to calculate a MET score were followed for dementia mortality (see study flow in Figure 7). Other reasons for termination of the follow-up were death for another reason, moving abroad, or the end of the follow-up period in the end of the year 2011. There were altogether 21,791 twin individuals who had answered the questionnaires adequately. Of these 21,791 persons, 267 had moved or died before their questionnaire answers were transferred into the data. Thus, there were altogether 21,524 persons in the follow-up. The number of twin individuals for whom information on all the covariates was also available was 20,404 (8988
Participants and Methods

pairs, 6455 MZ individuals, 13,765 DZ individuals, and 193 twin individuals with uncertain zygosity).

All twins aged at least 65 years have been invited to participate in health and cognition interviews (see study flow in Figure 7). In 1999–2001, MZ twins born in 1937 or earlier and with both co-twins alive were interviewed. In 2003–2007, DZ twins and twins of unknown zygosity born in 1937 or earlier were interviewed. Twins born in 1938–1944 were interviewed in 2014–2016. In Study II, the twins who had been interviewed and whose interview results had been entered into the data by 12 November 2015 were included in the study. The number of these twin individuals was 3050 (994 full twin pairs; 405 MZ, 570 DZ, and 19 pairs of unknown zygosity). The number of twins with full data on all the covariates was 2927 twin individuals. The participation rate in this study was 78% among those still alive and with an address in Finland.

MOBILETWIN is an ancillary study to the older Finnish Twin Cohort Study. For this ancillary study, twins born in 1940–1944 were selected. There were altogether 3186 twin individuals who belonged to these birth cohorts and had answered at least one of the questionnaires from 1975 and 1981. Other criteria for the invitation were that both members of the twin pair were alive and had a functional address in Finland. Of this cohort, 145 were excluded due to earlier participation in mental disorder studies (schizophrenia and bipolar disorder). There were altogether 1632 twin individuals who fit these criteria and had not participated in the mental disorder studies. 1012 (61.9%) twin individuals participated in the health and cognition interview. All twins who participated in the health and cognition interviews were offered the possibility to participate in one-week objective PA and SB monitoring with an accelerometer. In conjunction with the accelerometer monitoring, a questionnaire on physical functioning and mobility was administered. 791 twin individuals finished the accelerometer monitoring (54% of those alive and with an address in Finland) and 817 twin individuals completed the physical functioning and mobility questionnaire. The number of twin individuals who completed all three parts (health and cognition interview, accelerometer monitoring, and physical functioning and mobility questionnaire) was 785. The final analyses of Study III comprised of 726 twin individuals who also had full information on the covariates used. Most of the twins who participated were community-dwelling (opposed to living in an institution).
Same-sex twins born in Finland before 1958

Alive in 1967 (N=43 228)

- Exclusions and losses to follow-up (n=16 666)
  - deceased (1901)
  - failed to respond (7249)
  - lost address (2848)
  - excluded because not biological twin (4583)
  - declined (70)
  - other reasons (15)

Alive in 1981 (N=33 532)

- Exclusions and losses to follow-up (n=8878)
  - deceased (170)
  - failed to respond (4659)
  - lost address (1184)
  - excluded because not biological twin (28)
  - declined (1501)
  - other reasons (1 336)

Exclusions and losses to follow-up (n=16 666)

- deceased (1901)
- failed to respond (7249)
- lost address (2848)
- excluded because not biological twin (4583)
- declined (70)
- other reasons (15)

Exclusions and losses to follow-up (n=8878)

- deceased (170)
- failed to respond (4659)
- lost address (1184)
- excluded because not biological twin (28)
- declined (1501)
- other reasons (1 336)

21 524 twin individuals with answers to PA questions in both 1975 and 1981

Study I

- ≥ 65-year-old members of the Older Finnish Twin Cohort (born in 1937–1944 or earlier) were invited to participate in cognition interviews
  - both co-twins alive with known address in Finland
  - has not participated in selected earlier studies implemented

3050 twin individuals with complete data on PA from both questionnaires and cognition interviews completed by 12 November 2015 (994 twin pairs)

Study II

- 11 060 born in 1945 or later
- 4144 deceased
- 276 un-contactable
- 841 declined, or the spouse or guardian of the cohort member declined
- 22 failed to finish the cognition interviews
- 8 excluded because of contradictory data (mistake suspected)
- 2390 excluded for other reasons (mainly because interview not yet conducted)

Twins born in 1940–1944 who have filled in health questionnaires in 1975 or 1981 and possibly in 1990 (N=3186)

1632 twins invited to cognition study (816 pairs)
- Both co-twins alive with known address in Finland
- Has not participated in selected earlier studies implemented in the cohort

444 declined
- 140 un-contactable
- 12 were not able to participate in the cognition interview or the spouse or the guardian of the cohort member declined
- 18 deceased
- 6 other reasons

1012 twin individuals participated in the cognition interviews (61.9%)

- 127 did not want to participate in the accelerometer monitoring
- 100 with unsuccessful accelerometer data
- 6 individuals did not finish cognition interview but participated in the accelerometer monitoring

One-week accelerometer monitoring and PA questionnaire (for all those who participated in the cognition interviews):
- 785 individuals have participated in all three parts (cognition interview, activity questionnaire, and accelerometer studies) (including 281 complete pairs)
- 726 individuals with information on all of the covariates (including 235 complete pairs)

Study III

Figure 7. Study flow
4.2 MEASURES

4.2.1 PHYSICAL ACTIVITY
In Studies I and II, information on PA was derived from the questionnaires from 1975 and 1981. Both the intensity of PA and the volume of PA were assessed. Information on the intensity of PA was derived from the following question:

Is your physical activity during leisure-time about as strenuous on average as:
1) walking
2) alternately walking and jogging
3) jogging
4) running

Vigorous PA was defined as choosing one of the options 2–4 (Kujala et al., 1998).

Information on the volume of PA was based on the following questions:

How long does the physical activity last at one session on average:
 a) less than 15 minutes (class midpoint 7.5 min)
 b) 15 min – less than 30 min (22.5 min)
 c) 30 min – less than 1 hour (45 min)
 d) 1 hour – less than 2 hours (90 min)
 e) over two hours (120 min)

Presently how many times per month do you engage in physical activity during your leisure time:
 a) less than once a month (class midpoint 0.5)
 b) 1-2 times per month (1.5)
 c) 3-5 times per month (4)
 d) 6-10 times per month (8)
 e) 11-19 times per month (15)
 f) more than 20 times per month (20)

For the intensity of a session, the same question was used as for vigorous activity, but the following MET values were included for the choices:
 a) walking (corresponding to 4 METs)
 b) alternately walking and jogging (6 METs)
 c) jogging (10 METs)
 d) running (13 METs)
How much of your daily journey to work is spent in walking, cycling, running and/or cross-country skiing?

a) less than 15 min (class midpoint 7 min)
b) 15 min – less than a half an hour (22 min)
c) half an hour to less than an hour (45 min)
d) an hour or more (75 min)
e) I am presently not at work (0 min)

On the basis of these questions, the MET (metabolic equivalent that is a multiple of metabolic resting energy expenditure) score was calculated using the following formula: frequency x duration x intensity. The MET score was calculated for both LTPA and for PA during work journeys using the class midpoints, and these were subsequently added together (Kujala et al., 1998). The MET value 4 corresponding to walking was used in the calculation for PA during work journeys. Practically expressed, 1 MET h/day corresponds to 30 minutes of slow walking every other day. The MET score has been validated in the older cohort of the Finnish Twin Cohort Study (Waller et al., 2008).

1. To reflect the long-term level of PA (Kujala et al., 2002), compound variables combining the PA level from both questionnaires were created and used in the analyses. For the intensity of PA, three categories were created:

2. No vigorous PA in either of the years 1975 or 1981 (inactive according to long-term vigorous physical activity, i.e., VLinactive)
3. Vigorous PA in both years 1975 and 1981 (active according to long-term vigorous physical activity, i.e., VLactive)

Change (from no vigorous PA in 1975 to vigorous PA in 1981 or from vigorous PA in 1975 to no vigorous PA in 1981, change in long-term vigorous activity, i.e., VLchange)

The participants who dropped out of vigorous PA and who picked up vigorous PA between the years 1975 and 1981 were integrated to reach greater statistical power because no difference between these groups in preliminary analyses was seen.

1. For the long-term volume of PA, two different compound variables were created. First, the mean MET score from the questionnaires from 1975 and 1981 was calculated. Second, a compound variable to discern the most inactive individuals was created. The categorization was the following:

2. The participants who belonged to the most inactive MET quintile (I) in both years 1975 and 1981 (inactive according to long-term quantitative physical activity, i.e., QLinactive)
3. The participants who belonged to other quintiles (II–V) than the most inactive quintile (I) in both years 1975 and 1981 (*active according to long-term quantitative physical activity, i.e., QLactive*)

Change (from the most inactive quintile (I) in 1975 to other quintile (II–V) in 1981 or from an active quintile (II–V) in 1975 to the most inactive quintile (I) in 1981, *change in long-term quantitative activity, i.e., QLchange*)

In Study III, accelerometers were used in evaluating PA levels. The accelerometer type was the hip-worn Hookie AM20 (Traxmeet Ltd, Espoo) that uses a light digital triaxial acceleration sensor (ADXL345; Analog Services, Norwood, MA, USA). The device stored the raw acceleration signals with 100 Hz sampling frequency. The raw acceleration data was turned into the mean amplitude deviation (MAD) values independent of accelerometer brand (Vähä-Ypyä et al., 2015a). The MAD value describes the resultant acceleration during a 6-second (s) epoch. The MAD values are strongly correlated with oxygen consumption (VO2) (Vähä-Ypyä et al., 2015b) and can be converted into MET values. The posture in stationary activities was determined using the angle for posture estimation (APE) (Vähä-Ypyä et al., 2018).

The participants were instructed to wear the accelerometer during all waking hours but to remove the device when entering a sauna, shower, or swimming pool. The participant’s recording was included when it had at least 10 hours of data from at least 4 days. Time periods over 30 minutes without any acceleration signal have been interpreted as “non-wearing time”. On average, the accelerometer monitoring was done 3.5 weeks (standard deviation (SD) 5.5) after the cognition interviews. The dimensions of PA and SB examined in this study were: SB including sitting and lying but not standing (SB), LPA (1.5–3 METs), MVPA (over 3 METs), daily step count, and the mean daily MET. The mean daily MET is the average MET of all waking hours.

4.2.2 DEMENTIA MORTALITY AND COGNITION

Dementia mortality was followed for those participants who had answered both the questionnaire in 1975 and in 1981, providing the required information to calculate the MET score. The follow-up period ranged from 1981 to the end of the year 2011. The causes of death and exact dates of death and emigration were acquired from the Cause of Death Register. Dementia mortality was defined according to the contemporary ICD classification. The following categories were classified as dementia deaths: in ICD-8: category 290 (Senile and presenile organic psychotic conditions) and 794 (Senility without mention of psychosis); in ICD-9: category 290 (Senile and presenile organic psychotic conditions), 3310A (Alzheimer’s disease), 3311A (Pick’s disease), 4378A (Multi-infarct dementia), and 797 (Senility
without mention of psychosis); and in ICD-10: category F01 (Vascular dementia), F03 (Non-specific dementia), G30 (Alzheimer’s disease), G31.0 (Cerebral atrophy including Pick’s disease), and R54 (Senility). The contributing causes of death were not considered.

The telephone cognition screening included two telephone screening instruments: Telephone assessment of dementia (TELE) and Telephone Interview of Cognitive Status (TICS). The interviews were carried out by experienced trained research nurses and a PhD student from the Department of Public Health at the University of Turku. Notwithstanding the converging questions of TELE and TICS, the telephone cognition screening interview formed a 29-item linear variable representing global cognition (the total cognitive score) with higher points indicating better cognition.

The maximum scores for TELE and TICS are 20 and 33, respectively. Higher points indicate better cognitive function. The cut-off scores for cognitive impairment were <16 for TELE and <22.5 for TICS. The cut-off scores for healthy cognition were >17.5 for TELE and >26.5 for TICS (Laitala et al., 2008). The intermediate scores 16–17.5 for TELE and 22.5–26.5 for TICS and scores for those whose TELE and TICS scores are in disagreement were designated MCI (Vuoksimaa et al., 2016), although it has to be taken into account that our cognitive screening instruments are not validated for MCI.

4.2.3 COVARIATES

Data on basic demographics included age (at the time of the questionnaire in 1981 or at the time of the cognition interview), sex, and zygosity. For the vast majority, zygosity (MZ or DZ) had been ascertained with a validated questionnaire, but there were also some with unknown zygosity (incoherence in the validated questionnaire of zygosity).

In Study I and II, the additional covariates used were education, BMI, heavy drinking, hypertension, smoking, and for Study II, also the length of follow-up. Excluding the length of follow-up, all of these are self-reports, mainly from the 1981 questionnaire. Education was described in years of schooling ranging from 3 to 16 years (Silventoinen et al., 2000) and the information was based on the 1975 questionnaire if missing in the 1981 questionnaire. BMI was calculated from self-reported height and weight from the 1981 questionnaire. Heavy drinking was designated as consuming at least once a month and on a single occasion, more than five beers, a bottle of wine, or a half-bottle of spirits (Kaprio et al., 1987). This definition is very near the definition from the National Institutes of Health, defining binge drinking as drinking “5 or more (males) or 4 or more (females) drinks containing any kind of alcohol within a two-hour period” (National Institutes of Health, 2003). Hypertension was used as a binary variable based on questionnaire
Participants and Methods

information on reporting “permanently elevated blood pressure and medication for it” and on information on hypertension mediation provided by the Social Insurance Institution of Finland (Kujala et al., 2003). Smoking status was classified from a detailed smoking history questionnaire and used as a four-category variable (i.e., “never smoked”, “occasional smoker”, “former smoker”, “current smoker”) (Kaprio et al., 1988). For smoking status, also, the information on smoking was used from the 1975 questionnaire if it was absent in the 1981 questionnaire. In Study I, the length of follow-up was incorporated in the model (Cox regression analysis) and in Study II, it was calculated as the interval between the 1981 questionnaire and the cognition screening interview and used as a covariate in the regression model (logistic or linear regression model).

In Study III, the covariates were age, education, accelerometer wearing time, BMI, and living condition. The accelerometer wearing time was used as a covariate when studying SB and the mean daily MET. BMI was calculated from the height and weight reported in the questionnaire which was sent to the participants in conjunction with the accelerometer. Living condition was queried in conjunction with the cognition screening interview. The answers were dichotomized into two categories “living alone” or “living with a spouse, children or grandchildren, relatives, siblings, or other”.

In Studies I and II, the analyses were also done in a subgroup in which all the participants were healthy at baseline. These additional analyses were performed to eliminate chronic diseases’ effect on PA. The more detailed description on the exclusion for chronic diseases can be found in an earlier article from the research group (Kujala et al., 1998). In short, the most important chronic diseases including cancer, diabetes, chronic obstructive pulmonary disease, and cardiovascular diseases, notwithstanding hypertension or venous disease, were excluded based on information from the 1981 questionnaire, information on reimbursable medications from the Social Insurance Institution in Finland, and the data received from Finland’s nationwide hospital discharge register (between 1972 and 1982) and from the Finnish Cancer Registry (before 1983).

4.3 ETHICS OF THE STUDY

The study has been conducted according to good clinical and scientific practice and in accordance with the Declaration of Helsinki. The questionnaire studies in 1975 and 1981 have been approved by the National Board of Health of Finland and answering and returning the completed questionnaire was considered as consent. During the course of the study, the participants have been repeatedly informed about the study and they have been able to withdraw from it at any time if they so wished. The Ethics Committee of the Hospital District of Southwest Finland has approved the cognition
interview study and the extension with objective PA measurements. The twins aged at least 65 years were sent an invitation letter inquiring about their willingness to participate in the health and cognition interview and/or accelerometer study.

4.4 STATISTICAL METHODS

All analyses have been run with Stata (version 13 or 14, StataCorp LLC, College Station, TX, USA).

Two different analysis methods have been used: between-family (i.e., individual-level) analyses and within-family (pairwise) analyses. In between-family analyses, twins were treated as individuals, but because the members of a twin pair may be correlated, a robust cluster variance estimator was used (Williams, 2000). The second analysis method was within-family analyses. In these analyses, the members of a twin pair were compared with each other. That is to say that intrapair differences in PA levels were regressed on intrapair differences in cognition.

In Study I, a Cox regression model was used to estimate the association between PA levels and dementia mortality. In the within-family model of the Cox regression analysis, the data were stratified by pair, thus, providing us the within-pair estimates. In Study II and III, a linear regression model was used to estimate the association of PA levels and cognition. For between-family analyses, the robust cluster variance estimator was used to take into account the correlation between twin pair members. In within-family analyses, a fixed-effects model was used to investigate the within-pair differences. First, the command “xtset” was used to set Stata to handle the data in twin pair entities and then the fixed-effects model was used to provide the beta estimates for the within-pair differences. The command “dfadj” was used to correct for the standard errors by rectifying the degrees of freedom. The results are given as hazards ratios (HR) for the results from Study I and as beta estimates (β estimates) in Studies II and III. The statistical precision is shown with 95% CIs. In Study II, additionally, a multinomial logistic regression model was used to study ORs for cognitive impairment and mild cognitive impairment. For within-family analyses, a conditional logistic regression model was used. Additionally, the Sidak multiple comparison test was used to study differences in total cognitive score between different age groups in Study II. In Study III, Spearman’s correlation was used to examine the association of education and objectively measured PA level post hoc. Spearman’s correlation was used to exclude the effect of outliers.

The covariates were added to the model individually. Model 1 is the basic model adjusted for age and sex in Study I, for age, sex, and length of follow-up in study II and for age, sex, and accelerometer wearing time for the mean daily MET and SB in Study III. Model 2 is the final model with full adjustments of age, sex, length of education, BMI, hypertension, heavy drinking, and smoking in Study I, the same
adjustments plus the length of follow-up in Study II and adjustment for age, sex, length of education, BMI, living condition, and accelerometer wearing time for the mean daily MET and SB in Study III.

In Study I, the proportional hazard assumptions were assessed by visual inspection of the “log-log” curves by using stphplot except for the continuous variable (the mean MET score from 1975 and 1981) for which a test of the proportional hazards assumption was used. The test showed marginal significance at the 0.05 significance level. This significance can be considered minor in a cohort as large as ours. In Study II, the assumptions of the linear regression model and multinomial regression model were met for long-term vigorous activity groups and long-term quantitative activity groups. The mean MET score from 1975 and 1981 was not normally distributed and therefore a normally distributed log transformation of the mean MET score from 1975 and 1981 was used in the analyses. Because the results were fairly similar for the log transformed mean MET score from 1975 and 1981 and for the non-transformed mean MET score from 1975 and 1981, the results for non-transformed mean MET score from 1975 and 1981 are presented for the purpose of clarity. In Study III, the assumptions of the linear regression model were tested and met except for MVPA, for which the same procedure as for the mean MET score from 1975 and 1981 was applied: the analyses were also done with a normally distributed transformation of MVPA, but the results for the non-transformed MVPA are shown for the purpose of clarity.

The analyses were also done in Studies I–II for a subgroup of participants who were healthy in 1981 and for a subgroup who were younger than 60 years in 1981.

Interaction tests were used in all studies (I–III), and no interaction between PA levels by sex was found. The characteristics of the cohort are presented in means with SDs or ranges or in numbers with percentages. The group differences were analyzed with one-way ANOVA (for continuous variables) and with a likelihood ratio test (for categorical variables).
5 RESULTS

5.1 BASELINE CHARACTERISTICS

In Study I, the average age of the participants at baseline was 40.5 years (range 23.8–101.4 years), 52.5% of them were women and they had on average 7.9 years of education (see Table 6). The participants belonging to VLactive were younger, more educated, and had smaller BMIs than the participants belonging to the groups of VLinactive or VLchange (VLactive: age 35.2 years, education 8.9 years, and BMI 23.1 vs. for example VLinactive group: age 46.8 years, education 7.1 years, and BMI 24.6). A greater percentage of the participants in the VLactive group were men and heavy drinkers and the participants of VLactive group also exercised quantitatively more at baseline (MET 3.9 h/day vs. 1.4 h/day in the VLinactive group).

In Study II, the average age of the participants at baseline was 49.1 years (range 37.9–74.9 years), 49.1% of them were women, and their mean MET score from 1975 and 1981 was 2.3 h/day (see Table 6). The participants belonging to the group of VLactive were younger and their mean MET score from 1975 and 1981 was greater than for participants from the groups VLchange or VLinactive. The VLchange group was the most educated, had lower BMIs, and had the most heavy drinkers.

In Study III, the average age of the participants was 72.9 years (range 70.3–75.0 years) (see Table 6). They had on average 8.8 years of education and on average the time used in SB was 8 h 57 min during waking hours, in LPA 2 h 55 min, and in MVPA 39 min 55 sec; the average daily step count was 6383, and the mean daily MET 1.4 h/day (see Table 6–7). There was no significant difference in age, length of education, or the amount of SB and LPA for men and women. Men did, on average, significantly more MVPA and their mean daily MET and daily step counts were also significantly greater than women. All who finished the cognition interviews in birth cohorts 1940–1944 were invited to participate in the accelerometer monitoring. The number of twin individuals from the 1940–1944 cohort was 3186. The criteria for invitation into the cognition interviews was that both co-twins were alive, had a working address in Finland and had not participated in prior studies implemented in the cohort (schizophrenia, bipolar). The number of the invited twins was 1632 and, compared to the non-invited, they were younger, more educated, a greater percentage of them were women, and they had healthier lifestyle habits in midlife (fewer smokers and heavy drinkers, greater MET score). 1012 twin individuals participated in the cognition interviews and compared to those who declined the cognition interview, the participants were more educated, a greater percentage of
them were men, and fewer of them were smokers. No significant difference was found in age.

The accelerometer monitoring participants (791 twin individuals) compared to those who declined accelerometer monitoring but participated in the cognition interview (221 twin individuals) were more educated and fewer of them were heavy drinkers. No difference was found in age, sex, or midlife MET score, smoking, or percentage healthy.

Table 6 Baseline characteristics of the study cohorts

<table>
<thead>
<tr>
<th>Study</th>
<th>Age in years at baseline (mean, range)</th>
<th>Sex (% of women)</th>
<th>Education (years of schooling, mean, SD)</th>
<th>Mean MET from 1975 and 1981 (h/day, mean, SD)</th>
<th>Heavy drinking (n, %)</th>
<th>BMI in 1981 (mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study I (prospective)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (n=21 524)</td>
<td>40.5 (23.8–101.4)</td>
<td>11 306 (52.5)</td>
<td>7.9 (2.9)</td>
<td>2.4 (2.5)</td>
<td>5222 (24.8)</td>
<td>23.8 (3.4)</td>
</tr>
<tr>
<td>VL inactive</td>
<td>46.8 (23.9–91.1)***</td>
<td>5149 (61.7)***</td>
<td>7.0 (2.4)***</td>
<td>1.4 (1.3)***</td>
<td>1550 (19.3)***</td>
<td>24.6 (3.8)***</td>
</tr>
<tr>
<td>VL active</td>
<td>35.2 (23.8–80.8)</td>
<td>2779 (39.9)</td>
<td>8.9 (3.3)</td>
<td>3.9 (3.4)</td>
<td>2069 (29.9)</td>
<td>23.1 (2.8)</td>
</tr>
<tr>
<td>VL change</td>
<td>38.0 (23.8–101.4)</td>
<td>3378 (54.3)</td>
<td>7.9 (2.8)</td>
<td>2.1 (1.8)</td>
<td>1603 (26.2)</td>
<td>23.5 (3.3)</td>
</tr>
<tr>
<td>Study II (prospective)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (n=3050)</td>
<td>49.1 (37.9–74.9)</td>
<td>1496 (49.1)</td>
<td>7.6 (2.8)</td>
<td>2.3 (2.3)</td>
<td>462 (15.5)</td>
<td>24.9 (3.1)</td>
</tr>
<tr>
<td>VL inactive</td>
<td>51.4 (38.0–74.9)***</td>
<td>775 (60.5)***</td>
<td>7.0 (2.3)***</td>
<td>1.4 (1.2)***</td>
<td>152 (11.9)***</td>
<td>25.4 (3.3)***</td>
</tr>
<tr>
<td>VL active</td>
<td>46.7 (37.9–74.8)</td>
<td>434 (51.1)</td>
<td>7.6 (2.6)</td>
<td>3.6 (3.0)</td>
<td>124 (14.6)</td>
<td>24.9 (3.2)</td>
</tr>
<tr>
<td>VL change</td>
<td>48.3 (37.9–70.8)</td>
<td>287 (31.2)</td>
<td>8.5 (3.4)</td>
<td>2.2 (1.8)</td>
<td>186 (20.3)</td>
<td>24.3 (2.5)</td>
</tr>
<tr>
<td>Study III (cross-sectional)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (n=726)</td>
<td>72.9 (70.3–75.0)</td>
<td>352 (48.5)</td>
<td>8.9 (3.4)</td>
<td>8.8 (3.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>72.9 (71.2–75.0)</td>
<td>352 (48.5)</td>
<td>8.9 (3.4)</td>
<td>8.8 (3.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>73.0 (70.3–74.9)</td>
<td>374 (51.5)</td>
<td>8.6 (3.5)</td>
<td>8.8 (3.4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Study I and II, the groups of long-term vigorous activity were statistically significantly different for all baseline characteristics (oneway ANOVA, likelihood ratio test). In Study III, men and women did not differ significantly regarding age and the length of education (one-way ANOVA). The level of significance in differences between all groups in the row of the first group with the following marks: * p-value < 0.05, ** p-value < 0.01, ***p-value < 0.001.
Table 7 The physical activity and sedentary behavior profile in Study III

<table>
<thead>
<tr>
<th>Study III (cross-sectional)</th>
<th>SB (h:min, mean, SD)</th>
<th>LPA (h:min, mean, SD)</th>
<th>MVPA (mins, mean, SD)</th>
<th>Daily step count (mean, SD)</th>
<th>Mean daily MET (mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>8:57 (1:42)</td>
<td>2:55 (1:02)</td>
<td>39:55 (28:04)</td>
<td>6383 (3157)</td>
<td>1.4 (0.1)</td>
</tr>
<tr>
<td>Men</td>
<td>9:23 (1:39)</td>
<td>2:53 (1:00)</td>
<td>43:58 (30:12)**</td>
<td>6667 (3124)*</td>
<td>1.4 (0.2)**</td>
</tr>
<tr>
<td>Women</td>
<td>8:51 (1:44)</td>
<td>2:58 (1:35)</td>
<td>36:07 (25:22)</td>
<td>6116 (3168)</td>
<td>1.4 (0.1)</td>
</tr>
</tbody>
</table>

Men and women were statistically different for the amount of MVPA, the mean daily MET, and the daily step count (one-way ANOVA). The level of significance in differences between all groups in the row of the first group with the following marks: * p-value < 0.05, ** p-value < 0.01, ***p-value < 0.001.

5.2 SELF-REPORTED PHYSICAL ACTIVITY AND DEMENTIA MORTALITY

During the 29-year-long follow-up period, there were 5727 deaths and 353 had dementia as an underlying cause (Study I) in the cohort of 21,524 twin individuals (121 dementia deaths in MZ twins and 196 dementia deaths in DZ twins). Long-term vigorous activity was associated with a decreased HR for dementia mortality both in the model adjusted for age and sex (HR 0.65, 95% CI 0.43 – 0.98) and in the final model adjusted for age, sex, length of education, hypertension, smoking, BMI, and heavy drinking (HR 0.60, 95% CI 0.39 – 0.93) (Table 8). The mean MET score from 1975 and 1981 did not have a significant linear association with dementia mortality. The QLinactive group did not have a significantly elevated HR for dementia mortality when compared with the QLactive group and the QLchange group.

In the within-family analyses, the HR for the VLactive group was smaller both in Model 1 and in Model 2 but the results were not statistically significant (in Model 1: HR 0.48, 95% CI 0.17 – 1.32 and in Model 2: HR 0.39, 95% CI 0.12 – 1.26) (Table 9). There was no significant linear association between the mean MET from 1975 and 1981 and dementia mortality or between the groups of long-term quantitative activity and dementia mortality.

In both MZ and DZ twins, the VLactive group had a decreased HR for dementia mortality, but the results were not statistically significant in either MZ or DZ twins (HR for MZ twins in Model 2: 0.31, 95% CI 0.03 – 2.87 and HR for DZ twins in Model 2: 0.22, 95% CI 0.04 – 1.35) (Tables 10 and 11). The HR for dementia mortality for the VLactive group compared with the QLinactive group was smaller in MZ twins compared with DZ twins, but both results were statistically insignificant (HR for MZ twins in Model 2: 0.19, 95% CI 0.02 – 2.06 and for DZ twins: 0.65, 95% CI 0.10 – 4.28).
Results

There were altogether 20 twin pairs in which the twin pairs were discordant for dementia mortality and for long-term vigorous activity meaning that one twin member died of dementia and one did not and in addition one twin member belonged to the VLactive group and the other belonged to the VLinactive group. Of these twin pairs discordant for dementia mortality and long-term vigorous inactivity, 14 were DZ twin pairs and 6 were MZ twin pairs. In DZ twin pairs discordant for long-term vigorous activity, there were 8 pairs in which the inactive twin member died of dementia and 6 twin pairs in which the active twin died of dementia. In discordant MZ twin pairs, there were three pairs in which the inactive twin died of dementia and three twin pairs in which the active twin died of dementia. As for the long-term quantitative activity, there were 25 twin pairs in which the twin members were discordant for dementia mortality and one twin member belonged to the QLinactive group and the other twin member belonged to the QLactive group. Of these twin pairs discordant for QLactive and dementia mortality, there were 17 DZ twin pairs and 8 MZ twin pairs. Among the DZ twin pairs, there were 10 twin pairs in which the active twin died of dementia and 7 in which the inactive twin died of dementia. For MZ twin pairs, there were two twin pairs in which the active twin died of dementia and six twin pairs in which the inactive twin died of dementia.

When restricting the analyses to the participants who were healthy at baseline, the point estimates were very similar, but the CIs were wider and the results were not statistically significant. Another sub-analysis of participants 60 years old or younger was conducted and an association between long-term vigorous activity and decreased dementia mortality similar to the whole cohort analysis was observed (HR for dementia mortality for VLactive group compared with VLinactive group 0.49, 95% CI 0.26 – 0.94).
<table>
<thead>
<tr>
<th>Physical activity in 1975 and 1981 (questionnaire)</th>
<th>Dementia mortality (HR)</th>
<th>Total cognitive score (β-estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1 (age, sex)</td>
<td>Model 2 (age, sex, education, BMI, heavy drinking, hypertension, smoking)</td>
</tr>
<tr>
<td></td>
<td>n= 21,481</td>
<td>n=20,404</td>
</tr>
<tr>
<td><strong>Vigorous physical activity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLinactive</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VLactive</td>
<td>0.65 (0.43 – 0.98)</td>
<td>0.60 (0.39 – 0.93)</td>
</tr>
<tr>
<td>VLchange</td>
<td>0.95 (0.72 – 1.24)</td>
<td>0.92 (0.69 – 1.22)</td>
</tr>
<tr>
<td>Mean MET score from 1975 and 1981</td>
<td>0.98 (0.93 – 1.03)*</td>
<td>0.97 (0.92 – 1.03)*</td>
</tr>
<tr>
<td><strong>Volume of physical activity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QLinactive</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>QLactive</td>
<td>0.94 (0.70 – 1.27)*</td>
<td>0.85 (0.62 – 1.17)*</td>
</tr>
<tr>
<td>QLchange</td>
<td>1.17 (0.87 – 1.58)*</td>
<td>1.08 (0.79 – 1.48)*</td>
</tr>
<tr>
<td><strong>Physical activity in 2014–2016</strong></td>
<td></td>
<td>Model 1 (age, sex and accelerometer wearing time for the mean daily MET and SB) n=726</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=726</td>
</tr>
<tr>
<td>SB, h</td>
<td>-0.20 (-0.41 – 0.01)</td>
<td>-0.21 (-0.42 – (-0.003))</td>
</tr>
<tr>
<td>LPA, h</td>
<td>0.22 (-0.07 – 0.52)</td>
<td>0.30 (0.02 – 0.58)</td>
</tr>
<tr>
<td>MVPA, h</td>
<td>0.18 (-0.51 – 0.88)</td>
<td>-0.30 (-1.00 – 0.39)</td>
</tr>
<tr>
<td>Daily step count (thousands)</td>
<td>0.04 (-0.07 – 0.15)</td>
<td>-0.002 (-0.11 – 0.10)</td>
</tr>
<tr>
<td>Mean daily MET</td>
<td>0.94 (1.35 – 3.24)</td>
<td>-0.21 (-2.15 – 2.57)</td>
</tr>
</tbody>
</table>

(* Result not reported in original publications)
Table 9 Results from the within-family analyses for all twins

<table>
<thead>
<tr>
<th>Physical activity in 1975 and 1981 (questionnaire)</th>
<th>Dementia mortality (HR)</th>
<th>Total cognitive score (B-estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 (age) n=21,481</td>
<td>Model 2 (age, education, BMI, heavy drinking, hypertension, smoking) n=20,404</td>
<td>Model 1 (age, follow-up) n=3050</td>
</tr>
<tr>
<td>Vigorous physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLinactive</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VLactive</td>
<td>0.48 (0.17 - 1.32)</td>
<td>0.39 (0.12 - 1.26)</td>
</tr>
<tr>
<td>VLchange</td>
<td>0.97 (0.45 - 2.06)</td>
<td>0.95 (0.42 - 2.16)</td>
</tr>
<tr>
<td>Mean MET score from 1975 and 1981</td>
<td>0.91 (0.78 - 1.07)*</td>
<td>0.90 (0.76 - 1.07)*</td>
</tr>
<tr>
<td>Volume of physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QLinactive</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>QLactive</td>
<td>0.40 (0.14 - 1.13)*</td>
<td>0.34 (0.10 - 1.22)*</td>
</tr>
<tr>
<td>QLchange</td>
<td>0.65 (0.23 - 1.83)*</td>
<td>0.53 (0.17 - 1.65)*</td>
</tr>
<tr>
<td>Physical activity in 2014-2016 (accelerometer)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1 (age and accelerometer wearing time for the mean daily MET and SB) n=726</td>
<td>Model 2 (age, accelerometer wearing time for the mean daily MET and SB), BMI, living condition, education n=726</td>
<td></td>
</tr>
<tr>
<td>SB, h</td>
<td>-0.42 (-1.00 - 0.15)</td>
<td>-0.40 (-1.00 - 0.19)</td>
</tr>
<tr>
<td>LPA, h</td>
<td>0.60 (-0.21 - 1.42)</td>
<td>0.59 (-0.22 - 1.39)</td>
</tr>
<tr>
<td>MVPA, h</td>
<td>0.51 (-1.18 - 2.20)</td>
<td>0.20 (-1.64 - 2.04)</td>
</tr>
<tr>
<td>Daily step count (thousands)</td>
<td>1.11 (-0.36 - 0.38)</td>
<td>0.08 (-0.19 - 0.35)</td>
</tr>
<tr>
<td>Mean daily MET</td>
<td>3.20 (-2.59 - 8.98)</td>
<td>2.40 (-3.98 - 8.51)</td>
</tr>
</tbody>
</table>

(*Result not presented in original publications)
Table 10 Results from the within-family analyses for monozygotic twins

<table>
<thead>
<tr>
<th>Physical activity in 1975 and 1981 (questionnaire)</th>
<th>Dementia mortality (HR)</th>
<th>Total cognitive score (B-estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1 (age) n=6,364</td>
<td>Model 2 (age, education, BMI, heavy drinking, hypertension, smoking) n=6,068</td>
</tr>
<tr>
<td>Vigorous physical activity</td>
<td></td>
<td>Model 1 (age, follow-up) n=935</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model 2 (age, follow-up, education, BMI, heavy drinking, hypertension, smoking) n=896</td>
</tr>
<tr>
<td>VL inactive</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>VLA active</td>
<td>0.79 (0.14 – 4.45)*</td>
<td>0.31 (0.03 – 2.87)*</td>
</tr>
<tr>
<td>VLA change</td>
<td>1.13 (0.25 – 5.17)*</td>
<td>3.34 (0.34 – 33.2)*</td>
</tr>
<tr>
<td>Mean MET score from 1975 and 1981</td>
<td>0.98 (0.73 – 1.32)*</td>
<td>0.93 (0.67 – 1.29)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23 (-0.10 – 0.56)*</td>
</tr>
<tr>
<td>Volume of physical activity</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>QL inactive</td>
<td>0.18 (0.02 – 1.83)*</td>
<td>0.39 (0.02 – 2.06)*</td>
</tr>
<tr>
<td>QL active</td>
<td>0.20 (0.02 – 1.80)*</td>
<td>0.21 (0.02 – 1.96)*</td>
</tr>
</tbody>
</table>

Physical activity in 2014–2016 (accelerometer)

<table>
<thead>
<tr>
<th>Model 1 (age and accelerometer wearing time for the mean daily MET and SB) n=267</th>
<th>Model 2 (age, accelerometer wearing time for the mean daily MET and SB, BMI, living condition, education) n=267</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB, h</td>
<td>-0.74 (-1.63 – 0.15)</td>
</tr>
<tr>
<td>LPA, h</td>
<td>0.95 (-0.30 – 2.20)</td>
</tr>
<tr>
<td>MVPA, h</td>
<td>0.29 (-1.98 – 2.56)</td>
</tr>
<tr>
<td>Daily step count (thousands)</td>
<td>0.06 (-0.26 – 0.38)</td>
</tr>
<tr>
<td>Mean daily MET</td>
<td>4.60 (-3.39 – 12.6)</td>
</tr>
<tr>
<td></td>
<td>4.82 (-2.62 – 12.30)</td>
</tr>
</tbody>
</table>

(* Result not presented in original publications)
Table 11: Results from the within-family analyses for dizygotic twins

<table>
<thead>
<tr>
<th>Physical activity in 1975 and 1981 (questionnaire)</th>
<th>Dementia mortality (HR)</th>
<th>Total cognitive score (B-estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1 (age) n=13,564</td>
<td>Model 2 (age, education, BMI, heavy drinking, hypertension, smoking) n=12,874</td>
</tr>
<tr>
<td></td>
<td>Model 1 (age, follow-up) n= 2,049</td>
<td>Model 2 (age, follow-up, education, BMI, heavy drinking, hypertension, smoking) n=1967</td>
</tr>
<tr>
<td>Vigorous physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL inactive</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>VL active</td>
<td>0.34 (0.09 – 1.31)*</td>
<td>0.22 (0.04 – 1.35)*</td>
</tr>
<tr>
<td>VL change</td>
<td>0.81 (0.27 – 2.44)*</td>
<td>0.97 (0.27 – 3.53)*</td>
</tr>
<tr>
<td>Mean MET score from 1975 and 1981</td>
<td>0.78 (0.59 – 1.03)*</td>
<td>0.79 (0.58 – 1.09)*</td>
</tr>
<tr>
<td>Volume of physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QL inactive</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>QL active</td>
<td>0.56 (0.13 – 2.36)*</td>
<td>0.65 (0.10 – 4.28)*</td>
</tr>
<tr>
<td>QL change</td>
<td>1.00 (0.20 – 5.10)*</td>
<td>1.02 (0.14 – 7.48)*</td>
</tr>
<tr>
<td>Mean MET score from 2014 – 2016 (accelerometer)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB, h</td>
<td>-0.30 (-1.13 – 0.53)</td>
<td>-0.27 (-1.20 – 0.67)</td>
</tr>
<tr>
<td>LPA, h</td>
<td>0.48 (-0.64 – 1.60)</td>
<td>0.39 (-0.74 – 1.53)</td>
</tr>
<tr>
<td>MVPA, h</td>
<td>0.82 (-1.69 – 3.34)</td>
<td>0.56 (-2.35 – 3.46)</td>
</tr>
<tr>
<td>Daily step count (thousands)</td>
<td>0.12 (-0.31 – 0.57)</td>
<td>0.11 (-0.36 – 0.58)</td>
</tr>
<tr>
<td>Mean daily MET</td>
<td>3.35 (-5.21 – 11.9)</td>
<td>2.66 (-7.94 – 13.3)</td>
</tr>
</tbody>
</table>

(*Result not presented in original publications)
5.3 SELF-REPORTED PHYSICAL ACTIVITY AND COGNITION AFTER 25-YEAR FOLLOW-UP

After an average of 25.1-years follow-up (range 16.5 – 33.9 years), there were altogether 204 twin individuals with cognitive impairment. The beta estimates for total cognitive score according to the different levels of PA are shown in Tables 8 – 11. In between-family analyses, the participants belonging to the VL active group had significantly better cognitive function compared to the V Linactive in both the model adjusted for age, sex, and length of follow-up and the final fully adjusted model (beta estimate in Model 2: 0.91, 95% CI 0.47 – 1.35). The participants belonging to the VL change group had significantly better cognitive function than the V Linactive group in the age- and sex-adjusted model but not in the final fully adjusted model. The mean MET score from 1975 and 1981 was significantly associated with total cognitive score in the model adjusted for age and sex, but not in the fully adjusted model. The Q Linactive had significantly higher total cognitive scores than the QLinactive in both the model adjusted for age, sex, and length of follow-up and in the fully adjusted model (beta estimate in Model 2: 0.85, 95% CI 0.22 – 1.48).

The between-family analyses were also run in a subgroup of participants younger than 60 years at baseline. The results were very similar to those in the main analyses of the whole cohort. Long-term vigorous activity was associated with better total cognitive score in Model 2 (beta estimate 0.95, 95% CI 0.50 – 1.39) compared to the V Linactive group. The QL active group also had better total cognitive scores compared to the Q Linactive group (beta estimate in Model 2: 0.77, 95% CI 0.15 – 1.39). The mean MET score from 1975 and 1981 was significantly associated with total cognitive score in the model adjusted for age, sex, and length of follow-up, but not in the final fully adjusted model. In the subgroup of healthy participants at baseline, the results for long-term vigorous activity were very similar (VL active beta estimate: 0.84, 95% CI 0.36 – 1.33 in Model 2), but the significant association between long-term quantitative activity was only statistically significant in Model 1 and the mean MET score from 1975 and 1981 was not significantly associated with total cognitive score in either Model 1 or Model 2.

In within-family analyses, the only statistically significant result was that the group of VL change had significantly higher total cognitive scores than the V Linactive group in Model 1 but not in Model 2 (Table 11). For the VL active group, the point estimates were in the same direction as in the between-family analyses, but the point estimates were smaller and CIs were larger than in between-family analyses. The point estimate for the QL active group in the final fully adjusted model was similar as in the between-family models, but with wider CIs. The point estimate for the mean MET score from 1975 and 1981 was near to zero but in the opposite direction than in the between-family models.

In MZ twins, the beta estimates were in line with the beta estimates for all the twin pairs. The VL change group had significantly higher total cognitive scores than
the VLinactive group in Model 1 but not in Model 2. For the VLactive group, the beta estimates designated better cognitive function at the end of the follow-up, but the result did not reach statistical significance. In DZ twins, belonging to the VLactive or QLactive groups was associated with higher total cognitive scores, but CIs were large and not statistically significant.

In Table 12, the ORs for cognitive impairment and MCI are presented according to the different PA levels both from the between-family analyses and the within-family analyses. The VLactive group had significantly lower OR for cognitive impairment than the VLinactive group in both the model adjusted for age, sex, and length of follow-up and the final fully adjusted model (OR in Model 1: 0.55, 95% CI 0.45 – 0.67 and OR in Model 2: 0.78, 95% CI 0.63 – 0.96). The VLactive group also had a significantly decreased OR for MCI in Model 1 but not in Model 2 (OR in Model 2: 0.87, 95% CI 0.71 – 1.08). The VLchange group had a significantly decreased OR for both cognitive impairment and MCI in Model 1, but the results were not statistically significant in Model 2. In within-family analyses, the group of long-term vigorous activity had a decreased OR for cognitive impairment, but the point estimate was larger than in the between-family analyses, as were the CIs.

The ORs were also analyzed separately in a subgroup of participants who were healthy at baseline (Table 12). Long-term vigorous activity was associated with a decreased OR for both cognitive impairment and MCI in the model adjusted for age, sex, and length of follow-up but the association was not statistically significant in the final fully adjusted model. In within-family analyses, in the subgroup of healthy participants at baseline, the point estimates were similar as in the between-family analyses, but with larger and statistically non-significant CIs.

In Figure 8, the late-life total cognitive scores of different midlife PA levels are plotted by midlife age groups. It can be observed that the mean total cognitive score in the group of long-term vigorous activity deviates more clearly from the total cognitive score of the VLinactive group compared with the deviation observed in Figure 8B between QLactive and QLinactive groups. It can also be detected that the deviation is larger the older the age cohort is.

There were no twin pairs which were discordant for both baseline long-term vigorous activity (one twin in the VLactive group and the co-twin in the VLinactive group) and cognition (one twin with healthy cognition and the co-twin with cognitive impairment at the end of the follow-up). There were four twin pairs which were discordant for both baseline QLactive (one twin belonged to the QLactive group and the co-twin to the QLinactive group at baseline) and for cognition. Of these twin pairs, one was an MZ twin pair, and in this twin pair, it was the co-twin belonging to the QLinactive group who had cognitive impairment at the end of the follow-up. Three of these twin pairs discordant for baseline long-term quantitative activity and cognition were DZ, and in two of these pairs, it was the inactive co-twin who had cognitive impairment at the end of the follow-up. In one DZ twin pair, the active co-twin had cognitive impairment at the end of the follow-up.
<table>
<thead>
<tr>
<th>Long-term vigorous activity</th>
<th>Between-family analyses</th>
<th>Within-family analyses</th>
<th>Cognitive impairment</th>
<th>Mild cognitive impairment</th>
<th>Cognitive impairment</th>
<th>Mild cognitive impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive impairment</td>
<td>Model 1 (age, sex, follow-up)</td>
<td>Model 2 (age, sex, follow-up, education, BMI, heavy drinking, hypertension, smoking)</td>
<td>Model 1 (age, sex, follow-up)</td>
<td>Model 2 (age, sex, follow-up, education, BMI, heavy drinking, hypertension, smoking)</td>
<td>Model 1 (age, sex, follow-up)</td>
<td>Model 2 (age, sex, follow-up, education, BMI, heavy drinking, hypertension, smoking)</td>
</tr>
<tr>
<td>Mild cognitive impairment</td>
<td>Model 1 (age, sex, follow-up)</td>
<td>Model 2 (age, sex, follow-up, education, BMI, heavy drinking, hypertension, smoking)</td>
<td>Model 1 (age, sex, follow-up)</td>
<td>Model 2 (age, sex, follow-up, education, BMI, heavy drinking, hypertension, smoking)</td>
<td>Model 1 (age, sex, follow-up)</td>
<td>Model 2 (age, sex, follow-up, education, BMI, heavy drinking, hypertension, smoking)</td>
</tr>
<tr>
<td>All (n=3050 in Model 1 and n=2927 in Model 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL inactive</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>VL inactive</td>
<td>1.00</td>
</tr>
<tr>
<td>VL active</td>
<td>0.55 (0.45–0.67)</td>
<td>0.78 (0.63–0.96)</td>
<td>0.66 (0.54–0.81)</td>
<td>0.87 (0.71–1.08)</td>
<td>VL active</td>
<td>0.69 (0.44–1.09)</td>
</tr>
<tr>
<td>VL change</td>
<td>0.78 (0.65–0.94)</td>
<td>0.90 (0.74–1.10)</td>
<td>0.83 (0.68–1.00)</td>
<td>0.92 (0.76–1.12)</td>
<td>VL change</td>
<td>0.88 (0.59–1.34)</td>
</tr>
<tr>
<td>Healthy (n=2210 in Model 1 and n=2130 in Model 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL inactive</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>VL inactive</td>
<td>1.00</td>
</tr>
<tr>
<td>VL active</td>
<td>0.61 (0.48–0.77)</td>
<td>0.82 (0.64–1.05)</td>
<td>0.71 (0.56–0.90)</td>
<td>0.94 (0.73–1.21)</td>
<td>VL active</td>
<td>0.68 (0.38–1.21)</td>
</tr>
<tr>
<td>VL change</td>
<td>0.82 (0.66–1.03)</td>
<td>0.96 (0.76–1.21)</td>
<td>0.87 (0.69–1.09)</td>
<td>1.00 (0.79–1.27)</td>
<td>VL change</td>
<td>1.01 (0.57–1.78)</td>
</tr>
</tbody>
</table>

Table 12 Odds ratios for cognitive impairment and mild cognitive impairment according to long-term vigorous activity (Results not presented in original publications)
Results

A. Long-term vigorous activity

B. Long-term quantitative activity

Figure 8 Total cognitive score according to physical activity level in different baseline age groups (Results not presented in original publications)

*signifies a statistically significant difference in total cognitive score when compared to the VLinactive group in Figure 8A or to the QLinactive group in Figure 8B (the Sidak multiple-comparison test).

5.4 OBJECTIVELY MEASURED PHYSICAL ACTIVITY PROFILE AND COGNITION

In between-family analyses, there was no significant association between any of the variables describing PA or SB and total cognitive score in the models adjusted for age and sex (see Table 8). In the final fully adjusted model, LPA, and SB had significant associations with total cognitive score. LPA was positively associated with total cognitive score (beta estimate 0.30, 95% CI 0.02 – 0.58) and SB was inversely associated with total cognitive score (beta estimate -0.21, 95% CI -0.42 – (-0.003)). In post hoc analyses of the association of PA profile and length of education, MVPA was positively associated with the length of education and the results were statistically significant (Spearman’s correlation coefficient 0.12, p-value < 0.01). The Spearman’s correlation coefficient for LPA and the length of education was negative, and for SB and the length of education the Spearman’s correlation coefficient was positive (Spearman’s correlation coefficient for LPA -0.09, p-value 0.02; Spearman’s correlation coefficient for SB 0.02, p-value 0.67). For LPA and the length of education, the correlation was statistically significant.

In within-family analyses, the point estimates in the final fully adjusted models were quite close to those in between-family analyses but they did not reach statistical significance (see Table 8–9). The point estimates for MVPA, the daily step count, and the mean daily MET were statistically insignificant and slightly positive while in between-family analyses they were statistically insignificant and slightly negative.
In within-family analyses for MZ twins and DZ twins separately (see Tables 10–11), there were no statistically significant results, but point estimates for LPA and SB were similar as in between-family analyses and within-family analyses of all twin individuals.

Figure 9 shows the PA and SB profile according to cognition. The figure demonstrates how greater amounts of LPA and the smaller amount of SB are significantly associated with total cognitive score.

There was only one twin pair which was discordant for cognition: one twin member had healthy cognition and the other twin member had cognitive impairment. The co-twin with healthy cognition engaged more time in LPA and less in SB than the co-twin with cognitive impairment, but less time in MVPA than the co-twin with cognitive impairment. The daily step count and mean daily MET were fairly similar for these twins.

![Figure 9](image-url)

**Figure 9.** The average SB, LPA and MVPA amounts in hours per waking hours during one day according to different cognition groups (Healthy cognition: TELE score > 17.5 and TICS score > 26.5. Cognitive impairment TELE score < 16 and TICS score < 22.5. Mild cognitive impairment: TELE scores 16 – 17.5 and TICS scores 22.5 – 26.5 and those whose TELE and TICS scores are in disagreement.)
6  DISCUSSION

6.1  SUMMARY OF THE MAIN FINDINGS

The association of PA with dementia mortality and cognition was investigated in this thesis study (see Figure 10). Two prospective studies (I and II) investigated the association of self-reported PA with dementia mortality or cognition over two decades later. One cross-sectional study (III) addressed objectively measured PA and SB profile and cognition in late life. The studies were implemented in a twin study setting allowing for the possible confounding by genetic and shared childhood experiences to be taken into account. The members of a twin pair nearly always share their childhood environment. MZ twins’ DNA is identical at the sequence level and DZ twins share, on average, half of their segregating genes (like siblings).

At the individual level in these twins, long-term vigorous PA was associated with both decreased dementia mortality and better cognition over two decades later. The group of long-term quantitative activity (QLactive) also had significantly better cognition at follow-up than the QLinactive group after the lengthy follow-up, indicating that it is the most inactive persons who are at greatest risk of cognitive decline. MVPA in old age was not associated with better cognition, but LPA and smaller amounts of SB were at the individual level. In the within-family analyses in which the twin members are compared against each other, this thesis study showed no significant associations between PA level and dementia mortality or cognition in the final fully adjusted models, although the point estimates were quite similar to those in individual-level analyses. The number of twins discordant for PA and dementia mortality or cognition was very small in each original study. These facts imply that the associations found between PA levels and cognition at the individual level are most likely in part explained by genetics and shared childhood environment. In fact, a recently published GWAS found that a genetic variant in the APOE gene strongly implicated in AD is also associated with engagement in habitual PA (Klimentidis et al., 2018), providing preliminary evidence for genetic pleiotropy in the association of PA and cognitive decline, although in a surprising direction.

Further, the prospective follow-up study showed that PA had more influence on cognition the older the participants were at baseline indicating the danger of reverse causation when performing prospective studies of PA and dementia with baseline in old age. However, the significant associations at the individual level between long-term vigorous activity and long-term quantitative activity were also observed in this data in a subgroup of participants younger than 60 years at baseline.
### Intrinsic factors:
- **Age**: +++
- Genetics: +++ heritability for
  - AD: 33%
  - VaD: < 5%
  - Frontotemporal dementia: 50%
  - Lewy body dementia: 40%

### Extrinsic factors:
- Diabetes: +
- Hypertension: +++
- Obesity: +
- Physical inactivity:
  - Long-term vigorous physical activity is associated with decreased dementia mortality and better cognition over two decades later regardless of frequency
  - Belonging to the most inactive quintile in midlife entailed an increased risk for cognitive decline but not for dementia mortality
  - No dose-response was seen between PA and cognition or dementia mortality
  - Objectively measured late-life LPA was associated with better and SB with worse late-life cognition
  - In the analyses comparing twins with their co-twins, no statistically significant associations between PA and cognition or dementia mortality were seen despite the large twin cohort size
  - Very few twin pairs discordant for both physical activity and cognition or dementia mortality were seen suggesting that genetic selection or the effect of early childhood experiences affect the associations seen at individual level
- Smoking: +
- Education: +++
- Depression: +
- Head injuries: ++
- Hypercholesterolemia: +
- Alcohol consumption: +
- Sleep: +
- Stress: +

**Figure 10** How this thesis has contributed to the existing knowledge of the field. (Level of evidence: + low, ++ moderate, +++ robust) (Prince, M. et al., 2014, Plassman et al., 2010, Erkinjuntti et al., 2007)
6.2 SELF-REPORTED PHYSICAL ACTIVITY AND DEMENTIA MORTALITY

Long-term vigorous activity was associated with decreased dementia mortality at the individual level, but not significantly in analyses comparing twins against their co-twins. The volume of PA did not have a significant association with dementia mortality and belonging to the most inactive quintile both questionnaire years (1975 and 1981) did not increase the risk of dementia mortality significantly. These findings add to the existing knowledge of the field because there is a lack of prospective studies assessing midlife PA and its association with later dementia incidence. The earlier studies looking at vigorous PA specifically either found no significant association with dementia incidence (Gross et al., 2017) or found a significant association only in men (Tolppanen et al., 2015). This study result on long-term vigorous activity contradicts these findings. The association observed between long-term midlife vigorous PA and decreased dementia mortality at the individual level was found for both men and women, and no sex interaction was found. This finding on the volume of PA is in line with the study from Morgan et al. (2012), but conflicts with the findings from Gelber et al. (2012), Andel et al. (2008), and Chang et al. (2010) who either found an association for high PA volume and decreased dementia incidence (Gelber et al., 2012) or for regular or moderate PA and dementia incidence (Chang, M. et al., 2010, Andel et al., 2008).

This study had a similar result as two other prospective studies on PA and dementia mortality regarding vigorous PA (Rosness et al., 2014, Zotcheva et al., 2018). They are both from the Cohort of Norway (CONOR), but baseline information comes from different ages: either participants aged 50–74 years (Zotcheva et al., 2018) or 65–80 years (Rosness et al., 2014). The follow-ups lasted, on average, 10 (Rosness et al., 2014) or 15 (Zotcheva et al., 2018) years. The number of participants was large in both studies: 36,945 in that from Zotcheva et al. (2018) and 31,086 in that from Rosness et al. (2014). Both studies found that both light and vigorous PA are associated with decreased dementia mortality.

No significant association between long-term vigorous activity, long-term quantitative activity, or mean MET score from 1975 and 1981 with dementia mortality in analyses comparing twins against their co-twins was found. This is in line with the earlier twin studies in which no significant association between PA and dementia was found in co-twin control designs (Carlson et al., 2008, Andel et al., 2008). The number of twins discordant for both PA and dementia mortality was low. This study results, thus, suggest that the association found in individual-level analyses is likely to be due, to a moderate extent, to genetic selection or to childhood shared environment favoring both physically and cognitively healthy growth.
The meta-analyses on PA and dementia all report a significant association between PA and decreased incidence of dementia, but the association has been less clear when the baseline is in midlife and with longer follow-ups (Xu et al., 2017, Guure et al., 2017, Blondell et al., 2014, Hamer et al., 2009, Morgan et al., 2012). These meta-analyses have compared “high PA” with low or no PA, with “high PA” being heterogeneously described in individual studies. This study did not find a significant risk reduction of dementia mortality for high PA volume, but long-term vigorous PA regardless of the amount was associated with a decreased risk of dementia mortality. According to these results, it is specifically the ability and interest to participate in physical exercises more strenuous than walking that is influential.

The association of long-term vigorous activity and decreased dementia mortality was found in the model adjusted for age and sex and in the model taking the length of education, hypertension, smoking, BMI, and heavy drinking into account. The results do not seem, therefore, to be attributable to better cognitive reserve, reflected by the length of education, or to a healthier lifestyle. The analyses were also performed in a subgroup of healthy adults and the point estimates were similar to those in the whole cohort, but with wider and statistically non-significant CIs. This indicates that the results could be, to a moderate extent, explained by chronic diseases restricting one’s mobile ability or predisposing individuals to cognitive decline. This cohort included participants aged over 90 years at baseline, creating a possible risk of reverse causation in our study. Therefore, an additional analysis of participants aged 60 years at most was conducted. The VLactive group compared to the VLinactive group was also significantly associated with decreased dementia mortality in this subgroup, and the effect size was also considerable (HR 0.49, 95% CI 0.26 – 0.94).

Earlier, the meta-analysis of PA and dementia from Hamer et al. (2009) suggested differences by gender effects, but a more recent meta-analysis (Guure et al., 2017) on the subject showed very similar results for men and women. This study is in agreement with this because in the interaction test, no significant interaction by sex was found.

The results in the separate analyses of MZ and DZ twins are inconclusive because of the large spread of the data. Although the results were not statistically significant, the possibility of a type II error remains: we may falsely accept the null hypothesis of no significant difference between the dependent and independent variable.

6.3 SELF-REPORTED PHYSICAL ACTIVITY AND COGNITION

This study showed that long-term vigorous activity and long-term quantitative activity in midlife are significantly associated with better global cognitive function
Discussion

a quarter of a century later. This finding is in accordance with earlier high-quality studies (baseline in midlife, lengthy follow-up, at least moderately large cohort size > 1000, sound measures of both PA and cognition, and controlling for the most important confounding factors: age, sex, education, and vascular morbidity) reporting a positive association between PA and cognition (Singh-Manoux et al., 2005, Chang, M. et al., 2010, Virta et al., 2013a, Elwood et al., 2013). Singh-Manoux et al. (2005) did not address, however, global cognitive function and only one cognitive domain of four (crystallized intelligence) was significantly associated with midlife PA. Chang et al. (2010) also studied only specific domains, but in their study, all three domains (speed of processing, memory, and executive function) were positively associated with PA. Elwood et al. (2013) and Virta et al. (2013) also addressed global cognitive function. Elwood et al. (2013) reported that regular exercise was associated with better cognition, while the earlier twin study partially including the same study population as this thesis study (Virta et al., 2013a) did not show a significant association for vigorous PA and cognition in the fully adjusted models, but did show a significant protective association in the fully adjusted models for the highest PA quartile. This study adds to this earlier study (Virta et al., 2013a) in multiple ways: the number of twins with cognition screening is larger (3050 vs. 2165), PA has been studied more extensively, long-term engagement in PA is evaluated instead of a single measurement, and linear regression modeling has been used, providing more statistical power and enabling an assessment without the choice of categorization affecting the results.

It appears that it is particularly detrimental for one's cognitive health to be very inactive and to not engage in PA at all, because not belonging to the most inactive PA quintile in both 1975 and 1981 was associated with better global cognitive function. To be included in the most inactive quintile, the amount of PA was, at most, 0.6 MET-h/day, corresponding to 0.5–1 h of walking 3–5 times per month, hence strikingly little. In the subgroups of healthy participants at baseline, long-term vigorous activity but not long-term quantitative activity was associated with better cognition later. This may indicate that middle-aged patients with chronic diseases are especially in danger for developing cognitive decline because their chronic diseases limit their PA – whether it is by actual inability to participate in any kind of exercise or by inactivity brought on by psychological problems related to chronic diseases. The mean MET score from 1975 and 1981 was not significantly associated with cognition, indicating that there is no dose–response between the volume of PA and cognition.

The most recent meta-analyses have shown that PA is associated with a decreased risk of cognitive decline, but the association has been statistically significant only when the length of follow-up is fairly short (less than 5 or 10 years) (Morgan et al. 2012, Blondell et al., 2014) and when the baseline study population is past midlife (over 65 years) (Guure et al., 2017). Sofi et al. (2011) also found a significant
association between PA and better cognition for longer follow-ups (over 5 years), but they did not investigate the effect of baseline age on the results. These thesis results add to the previous knowledge showing a significant association between PA and cognition when baseline age is in midlife and the follow-up is considerably long. This thesis showed that midlife long-term vigorous activity and long-term quantitative activity were significantly associated with better cognition after a lengthy follow-up of over two decades. The results were significant even after controlling for age, sex, length of education, three vascular risk factors, and heavy drinking, suggesting no confounding by healthier lifestyle habits or higher cognitive reserve. The results for long-term vigorous activity and long-term quantitative activity were also significant in a subgroup of only participants younger than 60 years at baseline, and the result of long-term vigorous activity was also significant in a subgroup of healthy participants at baseline. As mentioned above, the number of methodologically sound, lengthy prospective studies assessing PA in midlife is not large and the meta-analyses might be limited by this fact.

The point estimates for long-term vigorous activity and long-term quantitative activity in the analyses comparing twins with their co-twins were very close to those in the individual-level analyses, but with wider and non-significant CIs. This finding implies that the association found at the individual level is, in part, explained by genetic selection and childhood shared environment. The number of twins discordant for PA and cognition was very small (zero for long-term vigorous activity and four for long-term quantitative activity). This fact strengthens the notion that the association found at the individual level is, to a quite moderate extent, explained by genetic selection and childhood shared environment. The earlier twin studies assessing PA and cognitive function and dementia (Andel et al., 2008, Carlson et al., 2008) did not find significant associations in their co-twin analyses comparing twins with their co-twins. Of these twin studies, Carlson et al. (2008) had the most dementia-discordant twin pairs (N=147), while the number of cognitively discordant twin pairs in our study was 120 and 90 in that from Andel et al. (2008), however, Carlson et al. (2008) used a fairly crude measure of PA. Yet again, the results in the separate analyses of MZ and DZ twins are inconclusive because of the large spread of the data. Although the results were not statistically significant, the possibility of a type II error remains: we may falsely accept the null hypothesis of no significant difference between the dependent and independent variable.

Causal relationships cannot be established in prospective cohort studies, but require evidence from RCTs and mechanistic studies. However, in longitudinal observational studies, a strong association and a dose–response effect between the independent variable and dependent variable can be considered an indication or reassurance of a possible causal relationship. No dose–response was observed between the volume of PA and cognition. It was specifically long-term vigorous activity and not belonging to the most inactive quintile in 1975 and 1981 that was
Discussion

associated with better cognition approximately 25 years later. However, recent studies suggest a U-shaped association between PA and cardiovascular health (Eijsvogels et al., 2018). Our preliminary scatter plots did not suggest a U-shaped association. The association of long-term vigorous PA and better cognition was not strikingly strong (beta estimate in Model 2: 0.91, 95% CI 0.47 – 1.35). This association strength does not confirm a causal relationship nor does it refute it.

Earlier studies have found an association between PA and cognition only in men (Tolppanen et al., 2015) or only in women (Kåreholt et al., 2011, Sumic et al., 2007), and in one meta-analysis differing effects on cognition in men and women were suspected to cause the heterogeneity observed in the results (Hamer et al., 2009). These thesis results are in agreement with the majority of studies reporting no differing effects for men and women.

The associations of long-term vigorous activity and cognition with multinomial logistic regression models were also investigated. The results essentially did not differ from the ones observed in the linear regression models. Long-term vigorous activity significantly reduced the risk of cognitive impairment in both the model adjusted for age, sex, and length of follow-up and the fully adjusted model. Long-term vigorous activity was significantly associated with decreased incidence of MCI in the model adjusted for age, sex, and length of follow-up, but not in the fully adjusted model. In within-family comparisons, no significant association was found although the point estimates were similar to those detected in the between-family analyses.

Our study participants’ age range at baseline was quite wide (38–75 years). In addition to the separate analysis in participants younger than 60 years, a figure of the association of long-term vigorous activity or long-term quantitative activity and cognition was drawn (Figure 8). The figure shows how the association of both measures of PA are more strongly associated with cognition at older ages expect for long-term quantitative activity in the oldest age group (n=76), most likely reflecting how reverse causation is a risk in dementia prevention studies. The results were, however, also significant in the subgroup of participants younger than 60 years at baseline.

6.4 OBJECTIVELY MEASURED PHYSICAL ACTIVITY AND SEDENTARY BEHAVIOR PROFILE AND COGNITION

In the cross-sectional study of Finnish elderly twins, objectively measured LPA was positively and SB inversely associated with cognition in individual-level analyses, but was not significant in analyses comparing twins with their co-twins. These findings indicate that the association could also be, in part, explained by genetic selection and childhood shared environment. This finding contradicts the trend observed in earlier cross-sectional studies of objectively measured PA profiles and cognition.
which implied that MVPA, but not LPA or SB, is associated with better cognition (Zhu, W. et al., 2015, Kerr et al., 2013). In our study, MVPA was not significantly associated with cognition. This is also interesting since it was specifically midlife vigorous PA that was associated with cognition in Studies I–II. However, the effect sizes for the association between LPA and SB with cognition in our study can be considered small (Cohen’s $d$ for LPA=0.08 and for SB=0.06). SB has not been associated with cognition in the earlier cross-sectional analyses of late-life objectively measured PA and cognition (Zhu, W. et al., 2015, Johnson, L. G. et al., 2016).

In cross-sectional studies addressing self-reported PA and cognition in older adults, the majority report a positive association. However, the influential aspects of PA and the affected domains of cognition vary (Bixby et al., 2007, Steinberg et al., 2015), but global cognition is also suggested to be affected (Lam et al., 2015). Nevertheless, contradicting evidence showing no association between self-reported PA and cognition in older adults also exists (Su et al., 2015). Gill et al. (2015) suggests that it is the life-long engagement in PA that has beneficial effects on cognition, and some studies suggest that it is the fitness level that is important for cognition (Zhu, N. et al., 2014, Scott et al., 2016). This thesis study adds to this developing heterogeneous field that with objectively measured PA, the effect size is not large at the individual level and may even be, in part, explained by genetic factors and childhood shared environment.

Earlier evidence from cross-sectional studies on the association of PA and cognition in young to middle-age is scarce and weak. A meta-analysis of cross-sectional studies on PA and cognition in young to middle-aged adults shows cautious evidence in favor of a positive association, but is limited by studies with mainly very small sample sizes (mean participant number when both active and inactive groups are included and when the single largest study is excluded: 59) (Cox et al., 2016)). The largest single study in this review showed an association between PA and working memory, but not with episodic memory, switching, or visuo-spatial memory (Erickson et al., 2013). Another singular study on PA and cognition found an association between PA intensity, but not duration, and overall cognition along with processing speed and mental flexibility (Angevaren et al., 2007). Longitudinal studies in children (Aggio et al., 2016) and in adolescence (Aaltonen et al., 2016) have shown that the direction of the association of PA and cognition seems to flow in the opposite direction: children and teenagers with better academic performance exercise, rather than the other way around. This cross-sectional study does not suggest any direction, but only an association in the elderly as well, regardless of direction or whether it is the recent months of exercise, the fitness level or the lifetime engagement in physical activities that is acting.

The fact that only associations of small effect size were found in this elderly cohort between PA and cognition is surprising because PA has been associated with significant improvement in cognition, especially global cognition, executive function,
and attention in elderly individuals with MCI (Öhman et al., 2014) and even in community-dwelling AD patients (Öhman et al., 2016, Vreugdenhil et al., 2012). Earlier studies have reported that master athletes perform cognitively superiorly to non-athlete controls (Tseng et al., 2013, Zhao et al., 2016). In this study, the average time engaged in vigorous intensity PA was 57 s and the maximum time engaged in vigorous intensity PA was 35.5 min. These small amounts of vigorous PA might not be enough to render cognitive benefits, but it has to be taken into account that in this age group the absolute PA intensity thresholds may not reflect the subjective PA intensities well (Kujala et al., 2017).

No significant associations between LPA and cognition or SB and cognition in the models adjusted for age, sex, and for the wearing time for SB were found, but the association was significant in the fully adjusted model. The individual covariate most affecting the significance of our finding was education. This is why the educational backgrounds according to PA levels were looked at. Post hoc the Spearman’s correlations between PA levels and education were investigated. LPA correlated negatively with education and the result was statistically significant. SB correlated positively with education, but the result was not statistically significant. This is surprising because in many study populations the situation is vice versa: better educated people engage more in PA and other healthier lifestyles. Because the associations between LPA and cognition and SB and cognition were very near the significance level in the models adjusted for age, sex, and the wearing time for SB (Model 1), the uneven distribution of education is likely to explain the different results between Model 1 and Model 2 (fully adjusted). The surprising distribution of education may be due to the exceptional post-war time period in Finland at the time our twin study cohort was young. Economic constraints in families may have prevented cognitively talented individuals from obtaining education or created pressure or compelling initiatives to earn money instead of acquiring mental capital. Very few Finns continued their formal education to the senior high school level (lukio in Finnish) in the pre- and immediate post-war years. Educational opportunities opened up to the whole population when comprehensive schooling for all children was enacted at the end of the 1960s (Anonymous, 1968).

Daily step count was not significantly associated with cognition. This is in line with our MET score results because walking at a pace greater than 3 km/h is already categorized as MVPA, and also MVPA was not significantly associated with cognition. Earlier studies using accelerometers did not assess the association of daily step count and cognition. The mean daily MET was also not significantly associated with cognition.

Again, the results in the analyses of MZ and DZ twins separately are inconclusive because of the large spread of the data. Although the results were not statistically significant, a possibility of type II error remains: we may falsely accept the null
hypothesis of no significant difference between the dependent and independent variable.

Our study population was slightly selected toward better educated individuals with healthier lifestyles (less smoking and heavy drinking) and most twins were community-dwelling. In an unselected study population with unhealthy lifestyle choices, lower educational levels and mobility limitations better represented, we might find more pronounced associations between PA levels and cognition. Mechanistic studies have shown that PA can ameliorate synaptic plasticity by strengthening synapses, inducing proliferation of dendrites, and boosting angiogenesis (Eadie et al., 2005, Lopez-Lopez et al., 2004), protect against vascular risk factors and poor sleep quality associated with cognitive decline, and even inhibit the accumulation of neuropsychological changes of AD (Adlard et al., 2005). In rodents, running in old age has also been associated with less accumulation of the plaques (Herring et al., 2016). Thus, a causal relationship between PA and cognition seems biologically plausible. However, in this cohort of 726 twins and 235 full twin pairs, only one twin pair was discordant for cognition. The weak association of objectively measured late-life PA and cognition could also be partly explained by genetic selection and childhood shared environment.

In summary, this thesis study indicates that the effect size of the association between late-life PA and late-life cognition is small and the association is at least partly explained by genetic selection and childhood shared environment.

6.5 STRENGTHS AND LIMITATIONS

This thesis study has several strengths. The study population in Studies I and II was mainly in their midlife and our follow-up in these studies was very extensive (29 years in Study I and 25.1 years in Study II). This kind of setting in a prospective study investigating degenerative diseases that develop very slowly is good. This thesis study of PA and cognition is, to my knowledge, the most extensive twin cohort study assessing the influence of PA on cognitive decline. The additional aspect of the co-twin control study design enables us to take genetics and childhood shared environment into account. This is especially valuable when studying traits that are heritable, as both PA and dementia or propensity for cognitive decline are.

We also had a fairly comprehensive set of information on other midlife lifestyle factors and some late-life lifestyle factors which enabled adjusting the results for many important possible confounding factors (midlife: age, sex, education, length of follow-up, BMI, smoking, heavy drinking, and hypertension and late-life: living condition, and BMI). Besides information from self-reports, we had well-documented information on chronic diseases from the nationwide hospital discharge register, from reimbursable medications from the Social Insurance Institution of Finland,
Discussion

and the Finnish Cancer Registry (Kujala et al., 2002), and separate sub-analyses in a subgroup of healthy participants at baseline in 1981 in Studies I and II were possible to conduct. An additional strength was that PA was evaluated twice, in 1975 and 1981, reflecting a long-term dedicated engagement in PA.

Because a very large proportion of the same-sexed twins born before 1965 with both co-twins alive in 1967 have participated in the studies, the study population can be considered a good representation of the Finnish general population. One of the strengths of the study is that in Finland we have numerous very reliable registers. In Study I, we used information from the Finnish Cause of Death Register. Also, the overall mortality of the older cohort of The Finnish Twin Study is similar to the general population (Kaprio, 2013).

This thesis study had also several limitations. The information on PA in Studies I and II came from questionnaire information. The self-reporting of habits is always liable for individual perception and interpretation. A recent large population-based study evaluated the concordance between PA questionnaire, diary, interview and accelerometer measurement in the same setting (Hukkanen et al., 2018). They found that PA estimates differ strongly between all methods except for the comparison between interview and questionnaire. Fitness level modulated the PA estimates: according to accelerometer monitorings, fitter persons underestimated their amount of MVPA in questionnaires. This is not surprising considering that MET levels measured with accelerometers are universal for all, although the relative intensity of exercise is subjective and varies greatly between fit and less fit persons (Kujala et al., 2017). A systematic review assessing PA questionnaires from 2012 concluded that while the reliability in the PA questionnaires evaluated was acceptable, the validity in them was moderate at best (Helmerhorst et al., 2012). The PA questionnaires of this review had been validated with accelerometers or with heart rate monitors or in rare cases with both of these. The MET index in this thesis study has been validated with a detailed PA telephone interview and the intraclass correlations were 0.68 (p < 0.001) for LTPA and 0.93 (p < 0.001) for commuting (Waller et al., 2008). The PA questionnaire in general has been considered valid in identifying between physically active and inactive people on the basis of association between PA levels and diseases and mortality in earlier studies (Kujala et al., 1998, Kaprio et al., 1978a, Kaprio et al., 1987, Kaprio et al., 1988, Kujala et al., 1994). From the era in question in Studies I-II at baseline (in the 1970s and 1980s), a validated structured questionnaire was the best available method to assess PA. Although, the absolute values for time used in PA may vary between methods used (Hukkanen et al., 2018), a questionnaire method still very likely adequately distinguishes between persons who are physically active and inactive.

Other midlife lifestyle variables were also self-reports, but many of these self-reports have been shown to be reliable and valid in Finnish population (regarding hypertension (Haapanen et al., 1997, Hernelahti et al., 1998), BMI (Mustelin et
al., 2009), and smoking (Vartiainen et al., 2002)). There are still confounding factors that we were unable to control for. Diet, for one, is a lifestyle factor that may influence risk of dementia (Eskelinen et al., 2011), and was not controlled for in this thesis study. Hypercholesterolemia was not added to the models because it was known only for a small proportion of study participants and its self-report has not been validated. It must be remembered that since the exact etiology of dementia is unknown, it is very challenging to study its risk and protective factors in a cohort design. Which covariates are confounding factors and which are mediators is not an easy task to deduce. This difficulty is emphasized in a study setting with a very extended follow-up during which the levels and engagement in independent factors and covariates may vary. The covariate choices in this thesis study reflect well the conception of the most important confounding factors of the author, but future research may yield a better and more precise understanding in this difficult matter. How to address possible confounders in a cohort study is always challenging, and robust causal interpretations cannot be made.

In Study I, dementia mortality as an outcome measure is crude and its sensitivity in detecting dementia cases is not optimal (Solomon, A. et al., 2014, Sommerlad et al., 2018). With a more sensitive outcome measure, more distinct associations between PA and dementia might be observed. The measure of dementia mortality in this thesis study is further limited by the fact that only underlying but not contributing causes of death were considered. It must also be noted, that the late rise of the incidence of dementia and cognitive impairment may introduce a survival bias. This means that persons developing dementia typically have to have lived long in order to develop dementia, creating a selected sample within the main cohort. Since we were not able to take competing events like coronary artery disease deaths or stroke into account, our results may be biased towards overly cautious estimates.

The outcome measure in Study II also has some limitations. Hearing problems can cause decreased cognitive scores in the telephone interview and the use of aids such as newspapers or calendars cannot be controlled. In the beginning of the telephone cognition interview, it was, however, asked not to use paper or pencil. Cognitive impairment is not the same as dementia and a diagnostic work-up including the thorough clinical examination, laboratory tests, broad neuropsychological battery of tests, or neuroimaging studies were not included in this thesis study. These kinds of extensive investigations would undeniably provide a more accurate estimate of dementia, but the telephone screening methods used, TELE and TICS, have been proven to be valid (Gatz et al., 2002, Brandt et al., 1998). The Finnish versions of them have also been validated and shown to correlate well with the results from the MMSE and to detect between dementia and healthy cognition with a sensitivity of 90.0% and specificity of 88.5% for TELE and a sensitivity of 86.7% and specificity of 88.5% for TICS (Järvenpää et al., 2002). In this thesis study, the cut-offs used for cognitive impairment and healthy cognition were even stricter and should, thus,
distinguish with even more accuracy between dementia and healthy cognition. Falling in between the two and having MCI according to the categorization in this thesis study has not been validated on the basis of these interviews.

An additional limitation in this thesis study is that cognitive impairment or general cognitive ability was not assessed at baseline. The majority of our study population in both prospective studies (I-II) were, however, in midlife, and with the long follow-up period, reverse causation is unlikely to affect the results much. The number of participants aged over 65 years at baseline in Study I was 1,356 out of 21,481 (6.3%) and in study II 76 out of 3050 (2.5%). Both studies were also analyzed in a subgroup of participants younger than 60 years and the results were very similar.

In Study III, PA was measured with a hip-worn accelerometer which can be considered objective and valid. One additional strength of the accelerometer is, that it can reliably measure the amount of SB. The amount of SB is more difficult to estimate in questionnaires than the amount of PA. However, accelerometer monitorings have their shortcomings, too. They do not record PA at the gym well and since our accelerometer was not water-resistant, all aquatic sports were also left uncaptured. However, the percentage of participants who engaged in aquatic sports during the accelerometer monitoring period was not high in this thesis study: 14.4%. The accelerometer method underestimates the PA intensity when skiing, bicycling, or doing forest work (wood chopping), but it does, however, sense these activities well enough to categorize them into MVPA. Because the preferred activity of the elderly is walking (Lim et al., 2005), the results of this thesis study have most likely captured the PA in this elderly study cohort of Study III quite well. The PA and SB profile of the study participants was comparable to that of the general population. In the study from Husu et al. (2016), the amount of LPA, MVPA, and SB were, in the elderly age group of persons aged from 70 to 85 years, approximately 2 h, 55 min and 8h, respectively, while in this thesis study time spent in LPA, MVPA, and SB were 2 h 55 min, 40 min, and 8 h 57 min, respectively.

Although, I consider that our protocol of adjusting with multiple other health indicators is well-based in its ability to eradicate the effect of other possible causal factors, it must be noted that adjusting for too many various factors predisposes the study to type II errors. The number of confounding factors used in this thesis study was moderate and, thus, I do not think a type II error in the main analyses of our study is probable. Additionally, the results were only adjusted for factors that are, according to our current understanding, important confounders and thus cannot be overlooked.

Because low birth weight may convey an increased risk for cognitive decline (Mosing et al., 2018) and for lower probability for undertaking LTPA (Andersen et al.) and twins are, on average, 600–1000 g lighter than singletons (Kyvik, 2000, Loos et al., 2005), the matter of generalizability has to be considered. However,
even if the risk of cognitive decline or dementia is elevated in this twin study, the estimates according to PA levels should not change because all of the cohort members were twins and, thus, the risk of cognitive decline might have been elevated in all of the subjects. In addition, twins catch up quickly in growth, showing only minor differences in anthropometric characteristics at the end of puberty. As mentioned above, it has been argued that twins are treated more similarly than non-twin siblings, creating a violation of one of the twin study assumptions (equal environments assumption) and correspondingly is counter-argued to be due to the likeness of the twins (Silventoinen et al., 2008). Another factor that limits the reliability of the co-twin control design is the fact that environmental and genetic factors most likely do not act completely separately of each other but that there are most likely genotype–environment interactions. This means that one type of genotype is more likely to end up in certain environments, creating a situation in which environmental effects have been affected by genetic factors. Although these interactions can be to some extent estimated in quantitative genetics, it is not possible to disentangle these in a traditional co-twin control design. This kind of violation of twin study assumptions would, however, only lead to an even greater effect of genetic influences, only emphasizing the main discovery of this thesis about genetics partly explaining the association between PA and cognition. As stated earlier in the literature review section, regarding LTPA, assortative mating has been shown not to occur in the Finnish population (Aarnio et al. 1997), but since genetic and environmental influences change between cultures and times, one study from the 1990s may not represent the current situation. On the other hand, assortative mating has been proven to occur for intelligence (Escorial et al., 2012), which may affect the susceptibility to develop dementia. Thus, although I consider this twin data as very reliable genetically informative data, the possible violations of these aforementioned assumptions of twin studies represent a small limitation in our study.

The study design in this thesis study was a co-twin control design. Thus, how large the proportion of the association is explained by genetics, shared environmental factors and unique environmental factors could not be evaluated. This is also a limitation of our study design. However, considering that only a handful of earlier studies on LTPA and cognition or dementia have been able to control for genetics and shared environmental factors in any fashion, the co-twin control design is primarily a strength.

The small number of discordant MZ twins has been considered a limitation in an earlier twin study of PA and cognition (Andel et al., 2008). In all the Studies I-III of this thesis, the number of twins discordant for PA and cognition or dementia mortality was very scarce. I argue that instead of reflecting small statistical power, this could be due to genetics partly explaining the association of PA and cognition or dementia mortality.
6.6 CONCLUSIONS

This thesis study showed that long-term vigorous PA is associated with decreased dementia mortality and better cognition over two decades later, but that genetic factors and childhood shared environment are likely to partly explain this association. This finding is new, and the subject has not been studied in a twin cohort as large as ours before. This thesis study implies that the strong association observed in earlier studies between PA and decreased incidence of dementia and cognitive decline could be partly explained by genetic selection. Recently, a genetic variant in the APOE gene strongly implicated in AD was associated with engagement in habitual PA, providing evidence of genetic pleiotropy in the association of PA and dementia (Klimentidis et al., 2018), although in a surprising direction. It must be remembered, though, that gene variants found in GWA studies have low penetrances.

The finding that midlife vigorous PA is associated with better cognition 25 years later adds to the scant body of truly long-term studies assessing midlife PA and late-life cognition. No significant association between the volume of PA and cognition approximately 25 years later was found. Long-term quantitative inactivity was significantly inversely associated with total cognitive score, but the result was not statistically significant in the subgroup of healthy participants at baseline. This implies that chronic diseases may partly explain the elevated risk of the most inactive participants.

Late-life LPA was positively, and late-life SB negatively, associated with cognition in this twin cohort. The associations were weak. Again, this thesis study showed that genetic factors and childhood shared environment could partly explain the association.

6.7 FUTURE PROSPECTS

This thesis study showed that at the individual level, midlife vigorous PA was associated with decreased dementia mortality and decreased incidence of cognitive decline over two decades later. A measure of midlife general cognitive ability was not available. In the future, we need more longitudinal studies that originate in midlife, before the long preclinical process of dementia begins, have long follow-ups and have also measured general cognitive ability in midlife. In a recent thesis study, it was shown that in old age, executive function predicts later life-space mobility when life-space mobility, a measure related to PA, did not predict later executive function (Poranen-Clark, 2018). In adolescents, academic performance predicted later PA, but PA did not predict later academic performance (Aaltonen et al., 2016). These mutual temporal associations from midlife to late life have not
been studied before to my knowledge. This kind of longitudinal study assessing physical functioning and cognition has, however, been published, and the study supported a direction from midlife cognition to later physical functioning rather than the opposite direction (Elovainio et al., 2009).

This thesis study showed that the association between midlife vigorous PA and decreased dementia mortality and better cognition is most likely explained partly by genetic selection. The co-twin control study design does not provide estimates on the extent to which the association is explained by genes and by the independent effect of PA. Complex human traits are seldom explained solely by nature (genes) or nurture (environment), and it seems probable the association observed at the individual level is a mixture of both effects: genes and the independent effect of PA. We need more genetically informative studies, both quantitative genetic modeling and multicenter twin studies with truly long-term follow-ups that originate in midlife and control for midlife general cognitive ability to clarify to what extent the association between PA and cognition is explained by genetic selection. The need for these kinds of studies is even greater since truly long-term RCTs with a large sample size and an efficient PA intervention with supervised PA sessions specifically studying the effect of PA on cognition are still lacking.

PA modifies the effect of genetic influences for type 2 diabetes (Kilpeläinen, 2009) and for obesity (Mustelin et al., 2009). To my knowledge, studies on whether PA modifies genetic influences on dementia or cognitive decline are scarce thus far. PA has been associated with both a more pronounced association with decreased incidence of dementia (Rovio et al., 2005), cognitive decline (Niti et al., 2008) or amyloid plaque accumulation measured with PET imaging (Head et al., 2012) in APOE ε4 carriers and with a more pronounced association with decreased dementia incidence in APOE ε4 noncarriers as well (Tolppanen et al., 2015, Podewils et al., 2005). In addition to clarifying PA's modifying effect on the association of the APOE ε4 allele with dementia, it would be very interesting to investigate whether PA modifies the effect of one's polygenic risk score (Desikan et al., 2017) on AD onset and the incidence of dementia and cognitive decline in general.

Understanding the complex association between cognition and LTPA throughout life is important to better target political resources and health education interventions where it really has most effect. Dementia is a progressive neurodegenerative disease leading to institutional care with high costs. Even small changes in lifestyles can produce large effects at the population level and lead to substantial savings in the health care system if independency in activities of daily living is preserved longer and institutionalization is prevented or postponed. To conclude, many aspects of the association between LTPA and cognition remain unsolved and would, through better scientific understanding, provide tools for better decision-making.

According to the literature I have read on the topic of physical activity and cognition and based on the results of this thesis study, I have to conclude that current
Discussion

evidence does not provide convincing evidence that physical activity’s power to prevent dementia would be large (Iso-Markku et al., 2016, Iso-Markku et al., 2015, Iso-Markku et al., 2018, Kivimäki et al., 2019, de Souto Barreto, Demougeot et al., 2018) and emphasis must rather be put on a balanced life combining many lifestyle strategies (Ngandu et al., 2015). I think that exercise may have some independent cognitive benefits and, according to this thesis, it is specifically the ability and interest to participate in physical exercises more strenuous than walking that is influential. This may also apply outside the realm of brain health. Over the past years, there has been a trend towards studying the beneficial health effects of LPA with encouraging results (Fuzeki et al., 2017). MVPA may, however, confer greater general health benefits (Saint-Maurice et al., 2018) and for some health effects, like arterial stiffness, LPA may not be a sufficient stimulus (Haapala et al., 2017b). In conclusion, my physical activity recommendation for both brain and general health would be to combine a balanced life with enjoyable exercise that causes sweating and some breathlessness in moderate amounts. Aristotle really made a good point, so long ago, with his lesson about the golden middle.
ACKNOWLEDGEMENTS

This work has been carried out at the Department of Public Health and the Department of Clinical Physiology and Nuclear Medicine, HUS Medical Imaging Center as part of the Doctoral Program in Population Health (DocPop), Doctoral School of Health, Faculty of Medicine, University of Helsinki in close collaboration with the Faculty of Sport and Health Sciences, University of Jyväskylä, Institute for Molecular Medicine Finland, University of Helsinki, the UKK Institute for Health Promotion Research, Tampere, the University of Turku, and the Clinical Neurology Department, Turku PET center, University of Turku. Present and former directors and department heads are respectfully thanked for providing excellent research facilities. In addition, this study has been financially supported by the Finnish Ministry of Education and Culture, Academy of Finland, Juho Vainio Foundation, HUS Medical Imaging Center, Finnish Medical Foundation, and by a special federal grant (EVO) to the Department of Clinical Physiology and Nuclear Medicine, a UH Medical Travel grant, and UH Chancellor’s Travel Grants, all of which I gratefully acknowledge.

I am extremely grateful for my two thesis supervisors, Dr. Katja Waller and Professor Urho Kujala. I sincerely thank Professor Urho Kujala and Dr. Katja Waller for the privilege of being taught and guided by them. Their vast knowledge of twin studies, the study of physical activity and research in general is very impressive. They always had time for my never-ending questions and they were very skilful in mending my mistakes in a diplomatic manner. I especially want to express my deepest gratitude to Katja for making my road less rocky with her caring guidance and to Urho for giving me the opportunity to pursue my dream of studying the health effects of physical activity and specifically the association of physical activity and dementia. To be able to combine research on physical activity with the wonder of the fleeting memory was the most interesting thesis topic I could have imagined. I want to you thank you both, Katja and Urho, for introducing me to the scientific world in such an inspiring and broad-minded way.

My deepest gratitude goes to Professor Jaakko Kaprio, the principal investigator of Finnish twin studies at the University of Helsinki, for giving me the opportunity to investigate my research questions with the excellent data from the older Finnish Twin Cohort Study and for all the wise comments which have not only improved my work but also developed me as a researcher.

Adjunct Professor Kari Kalliokoski and Docent Jenni Kulmala are truly appreciated for reviewing this thesis and I am honored to have Professor Kaare Christensen as my opponent in the public defence of this thesis.
Acknowledgements

I warmly thank Docent Eero Vuoksimaa for his expertise, valuable comments in the preparation of the manuscripts, and his aid in statistics, Kauko Heikkilä, the main database manager of the twin group, for his fast know-how and problem-solving with the data and my co-authors Professor Juha Rinne, Dr. Harri Sievänen, Noora Lindgrén, and Henri Vähä-Ypyä for their collaboration and valuable comments to improve the manuscripts.

I am grateful for all the twins who have donated their time and participated in making this study possible.

The Kaprio Group also deserves my thanks for welcoming me into their inspiring research group on the 7th floor during the last steps of my thesis journey. Your company created an excellently inspiring and enjoyable working environment and lightened the last steps of my thesis journey.

My sincere thanks to Alyce Whipp for the top-quality language review.

I am very grateful for my friends for counterbalancing the toils of work and Minna for providing excellent Word support even in challenging surroundings.

My warmest thanks go to my family: my mum, Liisa, dad, Jouko, and my two brothers Pentti and Pekka. First and foremost, I thank you for your never-ending love and support in all my endeavours. Without you, this thesis would not have been possible. And secondly, I thank you Pekka for your podcast recommendations which somehow ended up being pivotal in how I view statistics.
REFERENCES

Laki koulujärjestelmän perusteista 1968, .


References


Australian Imaging, Biomarkers and Lifestyle Study of Ageing”, *Molecular psychiatry*, vol. 18, no. 8, pp. 875-881.


de Souto Barreto, P., Andrieu, S., Rolland, Y., Vellas, B. & DSA MAPT Study Group 2018, “Physical activity domains and cognitive function over three years in older adults with...


References


Aerobic exercise improves hippocampal function and increases BDNF in the serum of young adult males", *Physiology & Behavior*, vol. 104, no. 5, pp. 934-941.


References


Kaprio, J. & Koskenvuo, M. 1988, “A prospective study of psychological and socioeconomic characteristics, health behavior and morbidity in cigarette smokers prior to quitting compared to persistent smokers and non-smokers”, *Journal of clinical epidemiology*, vol. 41, no. 2, pp. 139-150.

References


Kilpelaïnen, T. 2009, Physical Activity, Genetic Variation, and Type 2 Diabetes.


References


References


References


References


Peluso, M.A. & Guerra de Andrade, L.H. 2005, “Physical activity and mental health: the association between exercise and mood”, *Clinics (Sao Paulo, Brazil)*, vol. 60, no. 1, pp. 61-70.


References


Poranen-Clark, T. 2018, Relationship between cognitive performance and mobility over the lifecourse.


References


References


References


References


References


APPENDICES

Appendix I. Prospective studies assessing physical activity and cognition
Appendix II. Prospective studies assessing physical activity and dementia
Appendix III. Case-control studies assessing physical activity and cognition or dementia.
### Appendix I Prospective studies assessing physical activity and cognition

<table>
<thead>
<tr>
<th>Source/cohort (country)</th>
<th>Subjects (N= baseline target cohort, n= at follow-up or after declines), Sex</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Angevaren et al., 2010) (The Netherlands)</td>
<td>N=2434, n=1904 M, F</td>
<td>56</td>
<td>Total PA time (h/week) and average intensity (METS/week)</td>
<td>5</td>
</tr>
<tr>
<td>(Broe et al., 1998) (Australia)</td>
<td>N=613, n=327 M, F</td>
<td>80.6</td>
<td>Times per month 1) worked in garden or yard 2) did active sports or exercises, 3) went for walks</td>
<td>3</td>
</tr>
<tr>
<td>(Buchman et al., 2012) (USA)</td>
<td>N=1258, n=716 M, F</td>
<td>81.6</td>
<td>Total daily PA measured with 24h/d actigraphy for 10 days (in activity counts per day)</td>
<td>3.5</td>
</tr>
<tr>
<td>(Chang, M. et al., 2010) (Iceland)</td>
<td>N=5764, n=4945 M, F</td>
<td>51</td>
<td>Interview: 1) none 2) ≤ 5 hours/week 3) &gt; 5 hours/week</td>
<td>26</td>
</tr>
<tr>
<td>(Elwood et al., 2013) (UK)</td>
<td>n=2235, n=1225 M</td>
<td>45-59</td>
<td>Question: PA described as walking two or more miles to work each day, or cycling ten or more miles to work each day, or ‘vigorous’ exercise described as a regular habit;</td>
<td>30</td>
</tr>
<tr>
<td>(Etgen et al., 2010) (Germany)</td>
<td>N= 10 325, n= 3903 M, F</td>
<td>67.7</td>
<td>1) No activity (no regular PA) 2) Moderate activity (&lt; 3 times/week) 3) High activity (≥ 3 times /week)</td>
<td>2</td>
</tr>
<tr>
<td>(Flicker et al., 2005) (Australia)</td>
<td>N=12 203, n=618 M</td>
<td>77.4</td>
<td>Self-reported frequency of exercise (both vigorous and non-vigorous), (not specified further)</td>
<td>4.8</td>
</tr>
<tr>
<td>Assessment of cognitive function</td>
<td>Covariates</td>
<td>Main results (OR s or β-coefficients and 95% confidence intervals for cognitive decline, low or none PA used as a reference group if not mentioned otherwise)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Words Verbal Learning Test, the Stroop Colour Word Test, Letter Digit Substitution Test, Word Fluency Test</td>
<td>Baseline level of processing speed, age, gender, level of education, smoking, alcohol consumption, systolic blood pressure, total cholesterol, BMI, baseline level of PA (total time or intensity)</td>
<td>Beta coefficients for processing speed: Change in total PA time: 0.001 (-0.001, 0.002), change in average intensity: 0.063 (0.013, 0.113)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE, neuropsychological test battery: Logical Memory I and II, Visual Reproduction I and II subtests from the Wechsler Memory Scale-Revised, the Controlled Oral Word Association Test, the National Adult Reading Test, Cube Copying, Clock Drawing</td>
<td>Age, sex, years of education,</td>
<td>From vast test battery only walks were associated with only Logical Memory II (partial correlation -0,12, p-value &lt; 0.05) and Cube Copying (partial correlation -0,17, p-value &gt; 0.01). For MMSE: gardening -0,00 (p-value &gt; 0,05), active sports 0.06 (p-value &gt; 0,05), walks -0,01 (p-value &gt; 0,05).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 tests of cognitive function (not specified)</td>
<td>Age, sex, education</td>
<td>Rate of global cognitive decline: total daily PA estimate 0.033 (SE 0.012, p = 0.007) (linear mixed effects model)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive tests: 1) Speed of processing: digit symbol substitution test, Figure Comparison, modified Stroop test part I and II, memory: modified version of the California Verbal Learning Test, immediate and delayed recall, 3) executive function: Digits backwards, a shortened version of the CANTAB (Cambridge Neuropsychological Test Automated battery) Spatial Working Memory test, the Stroop test part III, 4) MMSE and DSM-IV diagnosis of dementia</td>
<td>Age, sex, education, body mass index, systolic blood pressure, smoking, cholesterol</td>
<td>β-coefficients: Speed of processing: medium: 0.22 (0.17–0.26), high: 0.32 (0.22–0.41), Memory: medium: 0.15 (0.10–0.20), high: 0.18 (0.07–0.29), Executive function: medium: 0.09 (0.10–0.14), high: 0.18 (0.09–0.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical records, Cambridge Cognitive Score, comprehensive diagnostics for those who screened positive in Cambridge Cognitive Score</td>
<td>Age, social class, National Reading Test</td>
<td>0.64 (0.41–0.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6CIT</td>
<td>Age, sex, BMI, baseline 6CIT score, depression, alcohol, diabetes, history of ischaemic heart disease and/or stroke, hyperlipidemia, hypertension, chronic kidney disease, smoking</td>
<td>OR: moderate: 0.57 (0.37–0.87), high: 0.54 (0.35–0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>Age, education, diabetes, drinking full cream milk, alcohol consumption</td>
<td>HR: 0.50 (0.25–0.99)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendices

<table>
<thead>
<tr>
<th>Source/cohort (country)</th>
<th>Subjects (N= baseline target cohort, n= at follow-up or after declines), Sex</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Gross et al., 2017) (USA)</td>
<td>N=1337, n=682, n=646 M, F</td>
<td>47 (38-63)</td>
<td>MET hours per day and engagement in regular exercise (At least once a week, do you engage in any regular exercise such as brisk walking, jogging, bicycling, etc., long enough to work up a sweat?)</td>
<td>30</td>
</tr>
<tr>
<td>(Ho et al., 2001) (Hong Kong)</td>
<td>N=2032, n=988 M, F</td>
<td>77.4</td>
<td>Exercise: yes or no</td>
<td>3</td>
</tr>
<tr>
<td>(Iwasa et al., 2012) (Japan)</td>
<td>N=1945, n=567</td>
<td>75.8</td>
<td>Engagement in regular physical activities (e.g. jogging, walking, Japanese croquet, hiking, dance, swimming, gymnastics): yes or no</td>
<td>5</td>
</tr>
<tr>
<td>(Jedrziewski et al., 2010) (USA)</td>
<td>N= 5280, n= 1203 (number of exercises) or 1240 (exercise sessions ≥ 20 minutes), M, F</td>
<td>65-110</td>
<td>1) Number of types of exercise performed 2) Number of exercise sessions ≥ 20 minutes</td>
<td>10</td>
</tr>
<tr>
<td>(Kim et al., 2011) (South Korea)</td>
<td>n=1204, n=518 M, F</td>
<td>Mean 72-78</td>
<td>PA: a self-report on 4-point scale: very active, fairly active, not very active, not at all active</td>
<td>2.4</td>
</tr>
<tr>
<td>(Kåreholt et al., 2011) (Sweden)</td>
<td>n=1643 M, F</td>
<td>46-75 (mean 57.4)</td>
<td>PA scale: doing sports, grading, dancing and no (0), yes sometimes (1), yes often (2) for each activity and a sum scale of these.</td>
<td>22.8</td>
</tr>
<tr>
<td>(Laurin et al., 2001) (Canada)</td>
<td>N=6434, n=4615 M, F</td>
<td>≥ 65</td>
<td>questionnaire 1) high ≥ 3x/week, at intensity greater than walking 2) moderate ≥3x/week, walking intensity 3) low: others 4) no PA</td>
<td>5</td>
</tr>
<tr>
<td>(Lee, S. et al., 2013) (Japan)</td>
<td>N not shown, n=550 M, F</td>
<td>≥ 60</td>
<td>Questionnaire: Light-intensity PA time per day (1.6–2.9 METs) and sedentary time (≤ 1.5 METs) (time, hours in both): quartiles in both</td>
<td>8</td>
</tr>
<tr>
<td>(Lytle et al., 2004) (USA)</td>
<td>N=1681, n=929 M, F</td>
<td>76.2</td>
<td>1) High exercise (aerobic exercise ≥ 30 minutes duration ≥ 3 times a week) 2) Low exercise (all other) 3) No exercise</td>
<td>2</td>
</tr>
</tbody>
</table>
### Assessment of cognitive function

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Main results (OR s or β-coefficients and 95% confidence intervals for cognitive decline, low or none PA used as a reference group if not mentioned otherwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>the Telephone Interview for Cognitive Status (TICS), animal naming, phonemic fluency (count of F, A, and S words produced), Hopkins Verbal Learning Test (HVLT), and the Brief Test of Attention. Average of z-scores of tests.</strong></td>
<td>Regular exercise: β-estimate for cognitive score Z-score units: 0.069 (95% CI: -0.092, 0.230) (n=285) β-estimate: -0.004 (95% CI: -0.231, 0.223) MET-hours per day: β-estimate: 0.012 (95% CI: 0.060, 0.084) (n=285) β-estimate: 0.039 (95% CI: -0.060, 0.138)</td>
</tr>
<tr>
<td>Age, sex, smoking, diabetes, and hypertension in 1978</td>
<td></td>
</tr>
<tr>
<td><strong>The information/orientation part of CAPE (the Clifton Assessment Procedure for the elderly)</strong></td>
<td></td>
</tr>
<tr>
<td>Age, sex, education, major income, residence, gait time, incident cerebrovascular disease</td>
<td>(No versus yes): OR 2.1 (1.3 -3.3)</td>
</tr>
<tr>
<td><strong>MMSE</strong></td>
<td>OR for cognitive decline for no participation in regular physical activities: 1.06 (0.65–1.74)</td>
</tr>
<tr>
<td>Age, gender, number of years of education, presence of chronic disease, IADL, depressive symptoms, smoking, hearing deficit, baseline MMSE</td>
<td></td>
</tr>
<tr>
<td><strong>Short Portable Mental Status Questionnaire (SPMSQ)</strong></td>
<td>Beta coefficient for Number of types of exercises performed and SPMSQ change (number of errors): -0.189 (p-value: 0.002) and beta coefficient for Number of exercise sessions ≥ 20 minutes and SPMSQ change: -0.045 (p-value: 0.007).</td>
</tr>
<tr>
<td>Age, sex, education, diabetes, hypertension, baseline score on cognitive test</td>
<td></td>
</tr>
<tr>
<td><strong>Screening: MMSE. Diagnoses: standard criteria for dementia.</strong></td>
<td>OR for cognitive decline (MMSE) (per descending category of activity i.e. very active, fairly active, not very active, not at all active): 1.51 (1.11–2.05)</td>
</tr>
<tr>
<td>Age, sex, education, depression, vascular risk scores, APOE genotype</td>
<td></td>
</tr>
<tr>
<td><strong>Items from MMSE</strong></td>
<td>Beta estimate for men -0.04, p-value: 0.607, beta estimate for women 0.17, p-value: 0.017</td>
</tr>
<tr>
<td>Age, age-square, follow-up-time, mobility problems, symptoms of mental distress, employment status, years of education, adult and childhood socioeconomic status, income, smoking, alcohol drinking</td>
<td></td>
</tr>
<tr>
<td><strong>3MS for screening, 3-stage clinical evaluation for diagnosis, NINCDS-ADRSA-criteria</strong></td>
<td>Cognitive decline: High: 0.58 (0.41–0.83)</td>
</tr>
<tr>
<td>Age, education, regular smoking, regular alcohol use, use of NSAID, functional ability in basic and instrumental activities of daily living, self-rated health, the number of chronic health conditions (only age and education for cognitive decline)</td>
<td></td>
</tr>
<tr>
<td><strong>MMSE</strong></td>
<td>OR for cognitive decline: light-intensity PA: lowest quartile: 1.00, second quartile: 0.58 (0.28–1.21), third quartile: 0.53 (0.25–1.12), highest quartile: 0.39 (0.19–0.83). OR for cognitive decline: lowest sedentary time quartile: 1.00, second quartile: 1.47 (0.74–2.89), third quartile: 1.37 (0.69–2.70), highest quartile: 2.66 (1.18–5.98)</td>
</tr>
<tr>
<td>Age, sex, educational level, smoking status, self-rated health, Center for Epidemiological Studies – Depression Scale, sleep duration, whether participant was working, hypertension, myocardial infarction, hyperlipidemia, diabetes mellitus, stroke, rheumatoid arthritis,</td>
<td></td>
</tr>
<tr>
<td><strong>MMSE</strong></td>
<td>OR: No exercise: 1.00, Low exercise: 0.69 (0.43–1.10), High exercise: 0.39 (0.19–0.78)</td>
</tr>
<tr>
<td>Age, sex, education, self-rating of health</td>
<td></td>
</tr>
<tr>
<td>Source/cohort (country)</td>
<td>Subjects (N= baseline target cohort, n= at follow-up or after declines), Sex</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Middleton, L. et al., 2008) (Canada)</td>
<td>N=6709, n=4683 M, F</td>
</tr>
<tr>
<td>(Middleton, L. E. et al., 2008) (Canada)</td>
<td>N=10263, n=5376 M, F</td>
</tr>
<tr>
<td>(Middleton, L. E. et al., 2011) (USA)</td>
<td>N=323, n=197 M, F</td>
</tr>
<tr>
<td>(Niti et al., 2008) (Singapore)</td>
<td>N=3577, n=1635 M, F</td>
</tr>
<tr>
<td>(Pignatti et al., 2002) (Italy)</td>
<td>n=364 F</td>
</tr>
<tr>
<td>(Pizzie et al., 2014) (USA)</td>
<td>N=173, n=91 M, F</td>
</tr>
<tr>
<td>(Psaltopoulou et al., 2008) (Greece)</td>
<td>N=1225, n=732 M, F</td>
</tr>
<tr>
<td>(Rajan et al., 2015) (USA)</td>
<td>N not shown, n=7742 M, F</td>
</tr>
<tr>
<td>(Richards et al., 2003) (UK)</td>
<td>N=5362, n=1919 M, F</td>
</tr>
</tbody>
</table>
### Assessment of cognitive function

#### Covariates
- Age, education, use of NSAIDs, sex, vascular risk factor index

#### Main results (ORs or β-coefficients and 95% confidence intervals for cognitive decline, low or none PA used as a reference group if not mentioned otherwise)

<table>
<thead>
<tr>
<th>Screening 3MS</th>
<th>Diagnoses: DSM-III-R (in addition: VCI-ND means in this article cognitive impairment, but not dementia &quot;of vascular origin&quot; i.e. having also history of cerebrovascular disease, stroke, atherosclerosis, or other cardiovascular disease, MCI means memory impairment, but no other cognitive impairment, intact instrumental activities of daily living, no diagnosis of dementia, other causes for CIND: depression, Parkinson etc.)</th>
<th>OR for CIND (moderate-high levels compared to low PA): 0.73 (0.59–0.91), for VCI-ND: 0.59 (0.40–0.88), for MCI: 0.96 (0.62–1.47). Significant sex modification of risk: for women CIND 0.62 (0.46–0.84) and for VCI-ND 0.34 (0.18–0.63), other (for MCI and all for men) non-significant and not shown.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3MS</td>
<td>None.</td>
<td>The proportion of high exercisers who had stable or improved cognition: 42.3% (40.6–44.0%), the proportion of low/no exercisers who had stable or improved cognition: 27.8% (26.4–29.3%). Khi2 test significant.</td>
</tr>
<tr>
<td>3MS</td>
<td>Baseline 3MS, age, sex, race, site, education, fat free mass, sleep duration, self-rated health, diabetes</td>
<td>OR for cognitive decline: Lowest tertile: 1.00, second tertile: 0.28 (0.06–1.23), highest tertile: 0.09 (0.01–0.79)</td>
</tr>
<tr>
<td>MMSE</td>
<td>Social activity, productive activity, age, gender, education, number of medical illnesses, vascular risk factors/events (hypertension, diabetes, cardiac diseases, stroke), smoking, alcohol drinking, physical functional status, depression, APOE ε4 status, baseline MMSE.</td>
<td>OR for cognitive decline: No: 1.00, Yes: 0.78 (0.60–1.02)</td>
</tr>
<tr>
<td>The Mental Status Questionnaire (MSQ)</td>
<td>Baseline MSQ</td>
<td>RR for cognitive decline for low PA: 3.7 (1.2–11.1)</td>
</tr>
<tr>
<td>Clinical Dementia Rating (CDR) based on Free and Cued Selective Reminding Test for verbal memory, the Trail Making Test Part A, Part B, category fluency and the Letter-Number Sequencing task.</td>
<td>Age at baseline PA assessment, gender, education, baseline cognitive performance on each test, APOE4 status, GDS (Geriatric Depression Scale-Short Form), parenta history of AD</td>
<td>No significant association with linear mixed effects modeling, exact results not shown (only a picture), but significant association for PA and cognition in at high risk patients</td>
</tr>
<tr>
<td>MMSE</td>
<td>Gender, age, marital status, years of education, height, BMI, smoking, alcohol intake, hypertension, diabetes, depression, daily energy intake, Mediterranean diet score</td>
<td>Multivariate linear regression analysis for the mean score of MMSE: Beta coefficient: 0.22 (0.02–0.42)</td>
</tr>
<tr>
<td>Battery of four tests: immediate and delayed recall of the East Boston Story, Symbol Digits Modalities test, MMSE. A composite score of these four tests.</td>
<td>Physical function, activities of daily living, depressive symptoms.</td>
<td>Linear mixed effects model: for white: PA &gt; 125 h/wk decreased cognitive decline with 20% (13–37%) (crude outcome, fully adjusted not shown)</td>
</tr>
<tr>
<td>15 item word learning task devised by the NSHD (the MRC (Medial Research Council) National Survey of Health and Development).</td>
<td>Spare-time activity, education, sex, manual social class, cognition at 15, physical disorder at 36, mental disorder at 36.</td>
<td>B estimate for change in verbal memory: Any PA 0.44 (0.01–0.87)</td>
</tr>
<tr>
<td>Source/cohort (country)</td>
<td>Subjects (N= baseline target cohort, n= at follow-up or after declines), Sex</td>
<td>Age at baseline</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>(Sabia et al., 2017) (UK)</td>
<td>n=10308, n=5123 M, F</td>
<td>44</td>
</tr>
<tr>
<td>(Schuit et al., 2001) (The Netherlands)</td>
<td>N=560, n=347 M</td>
<td>74,6</td>
</tr>
<tr>
<td>(Singh-Manoux et al., 2005) (UK)</td>
<td>N=10308, n=6236 M, F</td>
<td>35-55</td>
</tr>
<tr>
<td>(Sturman et al., 2005) (USA)</td>
<td>N=8501, n=4055 M, F</td>
<td>73,5</td>
</tr>
<tr>
<td>(Sumic et al., 2007) (USA)</td>
<td>n= 66 M, F</td>
<td>≥ 88.5</td>
</tr>
<tr>
<td>(van Gelder et al., 2004) (Finland, the Netherlands, Italy)</td>
<td>N=2285, n=295 M</td>
<td>Mean 73-76</td>
</tr>
<tr>
<td>(Wang, J. Y. et al., 2006) (China)</td>
<td>N=9341, n=5437 M, F</td>
<td>63</td>
</tr>
</tbody>
</table>
### Assessment of cognitive function

<table>
<thead>
<tr>
<th>Assessment of cognitive function</th>
<th>Covariates</th>
<th>Main results (OR or β-coefficients and 95% confidence intervals for cognitive decline, low or none PA used as a reference group if not mentioned otherwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive function: a measure of reasoning and 2 measures of verbal fluency. The Alice Heim 4-I (AH4-I) test was used to assess reasoning. Phonemic fluency was assessed via “S” words, and semantic fluency was assessed via “animal” words. Memory was assessed by using a test of short-term verbal memory that included a 20-word free recall test.</td>
<td>Age, sex, and socioeconomic position</td>
<td>0.95 (0.81–1.11)</td>
</tr>
<tr>
<td>MMSE</td>
<td>age, education, smoking, alcohol consumption, impaired cognitive functioning at baseline, medical history of myocardial infarction, angina pectoris, temporary ischaemic attack, non-insulin dependent diabetes mellitus, cerebrovascular disease.</td>
<td>OR for: ≤ 30 min/d: 2.0 (0.7–5.6), 31-60 min/d: 1.8 (0.8–5.1) and ≥ 60 min/d: 1.0</td>
</tr>
<tr>
<td>20-word free-recall test, Alice Heim 4-I measure of fluid intelligence, phonemic verbal fluency, semantic verbal fluency</td>
<td>Age, gender, education, socioeconomic position, self-rated health, blood pressure, cholesterol levels, smoking status, mental health status, social network, a proxy for baseline cognitive functioning (the Mill Hill Vocabulary Scale)</td>
<td>High PA from baseline used as a reference for low PA, ORs: memory: non-significant, Alice Heim 4-I: OR 1.65 (1.30–2.10), Phonemic Fluency: non-significant, Semantic Fluency: Non-significant</td>
</tr>
<tr>
<td>Global cognitive score from 4 cognition tests (East Boston Tests of Immediate Memory and Delayed Recall, MMSE, the Symbol Digit Modalities Test)</td>
<td>Age, sex, race, education, cognitive activities (model 2), depression, vascular diseases (model 3)</td>
<td>Mixed model beta coefficient in model 2: 0.0006 (p-value 0.10) and in model 3: 0.0005 (p-value 0.19)</td>
</tr>
<tr>
<td>MMSE, CDR (Clinical Dementia Rating Scale, biannual evaluation, low score on two consecutive assessments)</td>
<td>Age, education, apolipoprotein allele 4 status, and cognitive function (Delayed Recall Test)</td>
<td>Hazard ratio: for women 0.12 (0.03–0.41), for men 0.91 (0.25–3.40)</td>
</tr>
<tr>
<td>MMSE</td>
<td>Age, education, smoking status, alcohol consumption, country, mental activities, intensity/duration</td>
<td>Mean change in MMSE: ≤ 30 min/d: -1.1 (-2.3; 0.0), 31-60 min/d: -2.0 (-3.2; -0.8), 61-120 min/d: -1.2 (-2.0; -0.4), &gt; 120 min/d: -1.7 (-2.4; -0.9). Intensity: lowest quartile: -2.7 (-3.7; -1.8), second quartile: -1.1 (-2.0; -0.3), third quartile: -0.8 (-1.7; 0.1), highest quartile: -1.5 (-2.3; -0.6) (Khi2 tests)</td>
</tr>
<tr>
<td>The Chinese version of MMSE</td>
<td>age, sex, education, occupation, medical conditions, smoking, drinking, depressive symptoms, baseline MMSE, and ADL scores, and participation in other activities (cognitive or social).</td>
<td>All: 0.98 (0.95–1.01), Subjects with a full baseline MMSE score and did not develop cognitive impairment during the first year of follow-up: 0.96 (0.83–1.10) (n=1423)</td>
</tr>
</tbody>
</table>

**Note:** OR or β-coefficients (95% confidence intervals) for cognitive decline, low or none PA used as a reference group if not mentioned otherwise.
<table>
<thead>
<tr>
<th>Source/cohort (country)</th>
<th>Subjects (N= baseline target cohort, n= at follow-up or after declines), Sex</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Wang, S. et al., 2014) (USA)</td>
<td>n=1299, n=1249 F</td>
<td>83.3</td>
<td>Self-reported blocks walked per day and analyzed by tertile</td>
<td>5</td>
</tr>
<tr>
<td>(Vercambre et al., 2011) (USA)</td>
<td>N=3170, n=2809 F</td>
<td>Mean 72-73</td>
<td>Quintiles of total energy expenditure in exercise</td>
<td>5.4</td>
</tr>
<tr>
<td>(Verdelho et al., 2012) (Europe)</td>
<td>n=638, (51 failed to participate in the follow-up, they are assumed to be without cognitive decline) M, F</td>
<td>74.1</td>
<td>Interview: Physically active if at least 30 minutes of activity on at least 3 days per week.</td>
<td>3</td>
</tr>
<tr>
<td>(Vergheze et al., 2006) (USA)</td>
<td>N=488, n=353 M, F</td>
<td>Mean 79-80</td>
<td>PA score: Ten physical activities (tennis, golf, swimming, bicycling, dancing, group exercises, team games, walking, climbing more than two flights of stairs, babysitting). For each activity: 7 points for daily participation, 4 points for participating several days a week, 1 point for weekly participation, 0 points for participating occasionally or never.</td>
<td>5.6 (median)</td>
</tr>
<tr>
<td>Assessment of cognitive function</td>
<td>Covariates</td>
<td>Main results (OR s or β-coefficients and 95% confidence intervals for cognitive decline, low or none PA used as a reference group if not mentioned otherwise)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: 3MS, a measure of global cognition, digit span backwards, a test of working memory or attention, the California Verbal Learning Test II Short Form including a delayed recall portion, verbal fluency test (as many words as possible starting with the letter “f” in a minute, category fluency with naming as many vegetables as possible within a minute and Trails B. Diagnoses: DSM-IV (for dementia) and Petersen criteria (for MCI).</td>
<td>None.</td>
<td>No difference by activity level to the risk of MCI (22.8% for high, 24.0% for moderate, 22.2% for low, x2 test p = 0.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A global composite score (computed as the mean of the z-scores from all cognitive tests). Five cognitive tests: TICS, TICS 10-word list (immediate and delayed recalls), the East Boston Memory Test (immediate and delayed recalls, category fluency: naming as many animals as possible in one minute)</td>
<td>Age, education, marital status, alcohol intake, use of multivitamin supplements, smoking status, BMI, postmenopausal hormone therapy use</td>
<td>Mean differences in annual rates of cognitive change: reference: lowest quintile, second quintile: 0.00 (-0.01, 0.02), third quintile: 0.02 (0.00, 0.04), fourth quintile: 0.02 (0.001, 0.04), highest quintile: 0.3 (0.01, 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE, the Vascular Dementia Assessment Scale cognitive subscale (VADAS-Cog), Stroop and Trail Making tests as measures of executive function. NINCDS-ADRDA, NINDS-AIREN, McKhann GM, Albert MS, Grossman M, Miller B, Dickson D, Trojanowski JQ; Work Group on Frontotemporal Dementia and Pick's Disease. Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. Arch Neurol. 2001;58:1803–1809</td>
<td>Diabetes, global cognitive status at baseline</td>
<td>0.71 (0.257–0.940)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening and aMCI: the five-item Blessed-Test memory phrase. Diagnoses: DSM-III-R and subtyped using established criteria.</td>
<td>Age, sex, education, chronic illnesses</td>
<td>HR: 0.97 (0.933–1.008) (one-point increment PA score) for aMCI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source/cohort (country)</td>
<td>Subjects (N= baseline target cohort, n= at follow-up or after declines), Sex</td>
<td>Age at baseline</td>
<td>Assessment of PA</td>
<td>Follow-up (years)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>(Verghese et al., 2009) (USA)</td>
<td>N=488, n=401 M, F</td>
<td>Mean 79-80</td>
<td>PA score: Ten physical activities (tennis, golf, swimming, bicycling, dancing, group exercises, team games, walking, climbing more than two flights of stairs, babysitting). For each activity: 7 points for daily participation, 4 points for participating several days a week, 1 point for weekly participation, 0 points for participating occasionally or never.</td>
<td>4</td>
</tr>
<tr>
<td>(Weuve et al., 2004) (USA)</td>
<td>N=22 715 n=16 466 F</td>
<td>30-55</td>
<td>Quintiles of average MET from 8 to 15 years</td>
<td>1.8</td>
</tr>
<tr>
<td>(Willey et al., 2014) (USA)</td>
<td>N=3298, n=not shown M, F</td>
<td>69.2</td>
<td>MET-score 1) none 2) ≤ 14 3) &gt; 14</td>
<td>4.7</td>
</tr>
<tr>
<td>(Willey et al., 2016) (USA)</td>
<td>N=3298, n=993 M, F</td>
<td>71</td>
<td>Questionnaire, intensity, frequency and duration of various physical activities</td>
<td>5</td>
</tr>
<tr>
<td>(Virta et al., 2013a) (Finland)</td>
<td>N=2926, n=2 165 M, F</td>
<td>52</td>
<td>Questionnaire: 1) conditioner: ≥ 6x/month &amp; ≥ 30 minutes of exercise at least at the intensity of vigorous walking, 2) occasional: other PA, 3) no LTPA and MET quartiles in 1981</td>
<td>23</td>
</tr>
<tr>
<td>(Yaffe et al., 2001) (USA)</td>
<td>N=7701, n=5925 F</td>
<td>70-71</td>
<td>1) Quartiles according to blocks walked in a week 2) Quartiles according to total kilocalories expended per week</td>
<td>7.5</td>
</tr>
<tr>
<td>Assessment of cognitive function</td>
<td>Covariates</td>
<td>Main results (OR s or β-coefficients and 95% confidence intervals for cognitive decline, low or none PA used as a reference group if not mentioned otherwise)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnoses of dementia: DSM-III-R and subtyped using established criteria. MCI of vascular origin: (1) non-amnestic impairment: Digit Symbol Substitution, Digit span, Category Fluency test 2) cerebrovascular disease based on Hachinski Ischemic score symptom questionnaire, hemiparesis on clinical evaluation or history of stroke.</td>
<td>Sex, education, medical illnesses, the Blessed test score (baseline cognition) (age is the time scale in their statistical method)</td>
<td>HR for VCI: (1-point increment in Physical Activity Score): 0.993 (0.961–1.027)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TICS, East Boston Memory tests, category fluency, The Digit Span Backwards, delayed recall, global combined score</td>
<td>age, education, husband’s education, alcohol use, smoking status, aspirin use, ibuprofen use, vitamin E use, balance problems, health limitations in ability to walk a block, osteoarthritis, emphysema or chronic bronchitis, fatigue, poor mental health, antidepressant use, moderate to severe bodily pain</td>
<td>Mean differences (TICS): quintiles 1 (lowest) reference, quintile 2: 0.20 (0.07 to 0.32), quintile 3: 0.27 (0.15 to 0.40), quintile 4: 0.28 (0.15 to 0.40), quintile 5: 0.28 (0.21 to 0.47), (n=16 353) OR: 0.80 (0.67–0.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TICSm (modified TICS)</td>
<td>Age, ethnicity, sex, insurance, completing high school education, HDL, current tobacco use, moderate alcohol use, hypertension, current cholesterol lowering medications, depression, diabetes</td>
<td>Beta coefficient of MET-score (high vs. n PA): -0.03 (year) (p-value: 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuropsychological tests for episodic memory (list-learning: total recall, delayed recall, delayed recognition), executive function (time to complete the Color Trails test Form 1 and 2, Odd-Man-Out subtests 2 and 4 ), processing speed (eg. Groved Pegboard task), semantic memory (picture naming, category fluency and phonetic fluency)</td>
<td>Age, sex, education, Medicare status, crystallized abilities, time from baseline to first cognitive assessment, MVPA, smoking, alcohol consumption, BMI, hypertension, diabetes</td>
<td>Beta coefficient of LTPA (leisure-time PA: no or light vs moderate or heavy): for processing speed -0.23 (p=0.040), for executive function 0.005 (p=0.968), for semantic memory -0.045 (p=0.693), for episodic memory -0.223 (p=0.057).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TELE</td>
<td>Age, sex, education, the length of the follow-up</td>
<td>High PA used as a reference group: OR for sedentary: 2.52 (1.10–5.76), OR for lowest MET quartile 1.86 (1.27–2.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Mini-Mental State Examination (mMMSE)</td>
<td>Baseline age, educational level, health status, functional limitation, depression score, stroke, diabetes, hypertension, myocardial infarctation, smoking, estrogen use.</td>
<td>OR for quartiles according to walked blocks: lowest 1.00, second 0.87 (0.72–1.05), third 0.63 (0.52–0.77), highest 0.66 (0.54–0.82), OR for kilocalories per week: lowest 1.00, second 0.90 (0.74–1.09), third 0.78 (0.64 0.96), highest 0.74 (0.60–0.90).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendices

<table>
<thead>
<tr>
<th>Source/cohort (country)</th>
<th>Subjects (N= baseline target cohort, n= at follow-up or after declines), Sex</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Yaffe et al., 2009) (USA)</td>
<td>N=3075, n=2509 M, F</td>
<td>Mean 73-74</td>
<td>Weekly moderate or vigorous PA (e.g. aerobics, weight training or brisk walking at least once a week)</td>
<td>7</td>
</tr>
<tr>
<td>(Zhu, W. et al., 2017) (USA)</td>
<td>N=7098, n=6452 M, F</td>
<td>69.7</td>
<td>Accelerometer, moderate to vigorous PA quartiles</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Abbreviations: 6CIT = Six Item Cognitive Impairment Test, AH4-1 = the Alice Heim 4-1 test for logic skills, aMCI = amnestic mild cognitive impairment, CANTAB = Cambridge Neuropsychological Automated Battery, CAPE = The Clifton Assessment Procedure for the elderly, CDR = Clinical Dementia Rating, CIND = Cognitively impaired, no dementia, DSM-III-R = The diagnostic and Statistical Manual of Mental Disorders 3rd edition revised, HVLT = Hopkins Verbal Learning Test, MSQ = The Mental Status Questionnaire, NINDS-AIREN = The National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherché et l’Enseignement en Neurosciences, SPMSQ = Short Portable Mental Status Questionnaire, TICSm= modified TICS, VADAS-Cog = The Vascular Dementia Assessment Scale cognitive subscale
<table>
<thead>
<tr>
<th>Assessment of cognitive function</th>
<th>Covariates</th>
<th>Main results (ORs or β-coefficients and 95% confidence intervals for cognitive decline, low or none PA used as a reference group if not mentioned otherwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3MS</td>
<td>Age, ethnicity, education, REALM (Rapid Estimate of Adult Literacy in Medicine, word recognition test t test literacy), works or volunteers, caregiver (or not), social support, lives with someone, self-rated health, drinks &gt; 1 alcoholic drink per day, current smoking, depression, BMI, hypertension, diabetes, stroke/ TIA history, APOE ε4 allele, CRP, IL-6, triglycerides, fasting glucose</td>
<td>OR: moderate to vigorous exercise: (maintainer of cognition score vs. decliner): 1.31 (1.06–1.62)</td>
</tr>
<tr>
<td>Six Item Screener</td>
<td>Age, sex, race, region of residence, education, body mass index, hypertension, smoking and diabetes</td>
<td>OR for Quartile 2: 0.64 (0.48–0.84), Quartile 3: 0.53 (0.38–0.75), Quartile 4: 0.57 (0.39–0.82)</td>
</tr>
</tbody>
</table>

OR: odds ratio; β: beta coefficient; BMI: body mass index; CRP: C-reactive protein; IL-6: interleukin-6; TIA: transient ischemic attack; PA: physical activity; 3MS: three-minute swim test.
### Appendix II: Prospective studies assessing physical activity and dementia

<table>
<thead>
<tr>
<th>Source/cohort</th>
<th>Subjects at baseline (N), subjects at follow-up (n), Sex (M for men and F for women)</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (mean in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Abbott et al., 2004) (Hawaii, USA)</td>
<td>N=3734, n=2257 M</td>
<td>Mean age 76–77</td>
<td>Distance walked per day: &gt; 0.25 mile, 0.25 to 1 mile, &gt;1 to 2 mile</td>
<td>4.7</td>
</tr>
<tr>
<td>(Simons et al., 2006) (Australia)</td>
<td>N=3872, n=2805 (at baseline), n for follow-up not told M, F</td>
<td>≥ 60</td>
<td>Gardening (daily vs. rarely), Walking (daily vs. rarely)</td>
<td>16</td>
</tr>
<tr>
<td>(Akbaraly et al., 2009) (France)</td>
<td>N=7051, n=5698 M, F</td>
<td>Mean age 74–78</td>
<td>Score according to duration and activity (doing odd jobs, gardening, going for a walk): 0 (none ot less than 1 hour per day), 1 (1-2h per day), 2 (&gt;2 h per day). Range (0-6); low &lt; 2, mild = 2, high &gt; 2.</td>
<td>4</td>
</tr>
<tr>
<td>(Bowen, 2012) (USA)</td>
<td>N=856, n=808 M, F</td>
<td>Mean age 75–81</td>
<td>Vigorous physical score (on average engagement in activities such as aerobics, sports, running, swimming, bicycling, heavy housework, a job involving heavy physical work at least three times a week during the last 12 months): 0 (none) – 3 (vigorous activity at three follow-ups)</td>
<td>Median 5</td>
</tr>
<tr>
<td>(Broe et al., 1998) (Australia)</td>
<td>N=613, n=449 M, F</td>
<td>80.6</td>
<td>Times per month 1) worked in garden or yard 2) did active sports or exercises, 3) went for walks</td>
<td>3</td>
</tr>
<tr>
<td>(Buchman et al., 2012) (USA)</td>
<td>N=1258, n=716 M, F</td>
<td>81.6</td>
<td>Total daily PA measured with 24h/d actigraphy for 10 days (in activity counts per day)</td>
<td>3.5</td>
</tr>
<tr>
<td>(Carlson et al., 2008) (USA)</td>
<td>n= 147 twin pairs M</td>
<td>45</td>
<td>Participation in “outdoor activities”, “sports”, “gardening and home improvement” and “physical exercise after age 35” and a summary score of them (max. 4)</td>
<td>28</td>
</tr>
<tr>
<td>Assessment of dementia</td>
<td>Covariates</td>
<td>Main results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Screening:</strong> Cognitive Ability Screening Instrument (CASI), Diagnoses DSM-III-R, NINCDS-ADRDA, the criteria for California Alzheimer’s Disease Diagnostic and Treatment Centers for VaD.</td>
<td>Age, presence of apolipoprotein ε4 alleles, baseline Cognitive Abilities Screening Instrument score, decline in PA since mid-adulthood, physical performance score, years of education, BMI, childhood years spent living in Japan, status as a skilled professional, hypertension, diabetes, prevalent coronary heart disease, total and high-density lipoprotein cholesterol</td>
<td>Proportional hazard regression: Relative Hazard (RH) for &lt; 0.25 mile: 1.93 (1.11–3.34), 0.25 to 1 mile: 1.75 (1.03–2.99), &gt;1 to 2 mile: 1.33 (0.73–2.45), &gt; 2 mile: 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosis of dementia from medical records, if the participant has been placed in a nursing home</strong></td>
<td>Age, alcohol intake, gardening/walking, peak expiratory flow, depression score, marital status, education, prior history of stroke, activities of daily living</td>
<td>HR for dementia: gardening (daily vs. rarely): 0.64 (0.50–0.83), walking (daily vs. rarely): 1.00 (0.78–1.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-IV, NINCDS-ADRDA, NINDS-AIREN (National Institute of Neurological Disorders and Stroke-Association Internationale pour la Recherche en l’Enseignement en Neurosciences) criteria</td>
<td>Gender, educational level, occupational grade, study center, marital status, hypertension, diabetes, vascular diseases history, hypercholesterolemia, depressive symptoms, APOE genotype, incapacity in daily life activity, cognitive impairment assessed by MMSE (no mention when: at baseline or at follow-up)</td>
<td>HR: low: 1.00, mild 0.91 (0.59–1.39), high 1.09 (0.73–1.63) for all dementia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A battery of neuropsychological measures and standardized neurological examination</td>
<td>Age, sex, ethnicity, education, APOE genotype, smoking, alcohol consumption, BMI, diabetes, hypertension, stroke, heart disease, baseline TICS</td>
<td>OR 0.79 (0.64–0.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE, neuropsychological test battery, Logical Memory I and II, Visual Reproduction I and II subtests from the Wechsler Memory Scale-Revised, the Controlled Oral Word Association Test, the National Adult Reading Test, Cube Copying, Clock Drawing</td>
<td>Age, sex, education</td>
<td>Logistic regression analysis: no significant results (results not shown)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 tests of cognitive function (not specified), diagnoses, diagnoses NINCDS-ADRDA</td>
<td>Age, sex, education, BMI, depressive symptoms, vascular risk factors, vascular diseases.</td>
<td>HR 0.435 (0.244–0.778)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: TICSm (modified TICS) or IQCODE (Informant questionnaire on Cognitive Decline in the Elderly) if TICSm not possible. Diagnosis: neurological and neuropsychological evaluation including CERAD, final diagnoses by a consensus panel of neurologists.</td>
<td>None</td>
<td>OR: 0.99 (0.73–1.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source/cohort</td>
<td>Subjects at baseline (N), subjects at follow-up (n), Sex (M for men and F for women)</td>
<td>Age at baseline</td>
<td>Assessment of PA</td>
<td>Follow-up (mean in years)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
</tbody>
</table>
| (Chang, M. et al., 2010) (Iceland) | N=5764, n=4945  
M, F | 51 | Interview: 1) none ≤ 5 hours/week 3) > 5 hours/week                               | 26 |
| (de Bruijn et al., 2013) (The Netherlands) | N=10274, n=4406  
M, F | 72.7 | Log transformed MET hours per week                                                | 8.8 |
| (Elwood et al., 2013) (UK) | n=2235, n=1225  
M | 45–59 | Question: PA described as walking two or more miles to work each day, or cycling ten or more miles to work each day, or ‘vigorous’ exercise described as a regular habit; | 30 |
| (Gross et al., 2017) (USA) | N=1337, n=682, n=646  
M, F | 47 (38-63) | MET hours per day and engagement in regular exercise (At least once a week, do you engage in any regular exercise such as brisk walking, jogging, bicycling, etc., long enough to work up a sweat?) | 30 |
| (Gureje et al., 2011) (Nigeria) | N=2149, n=1225  
M, F | 74.5 | Low, moderate and high levels of PA (based on the International Physical Activity Questionnaire and categorized according to standard scoring criteria www.ipaq.ki.se) | 3 |
| (Hebert et al., 2000) (Canada) | N=8623, n=5747  
M, F | Mean age 76-78 | Engagement in regular exercise (not specified)                                   | 5 |
| (Karp et al., 2006) (USA) | N=2368, n=1810, n=1473, n=1375, n=732  
M, F | ≥75 | Interview: Physical component score; physical component score was assigned to each of the 29 activities and grading was based on authors’ own evaluations (low, moderate, high). Then the number of moderate/highly scored activities were summed up. | 6 (at 3 years demented were excluded) |
| (Kim et al., 2011) (South Korea) | n=1204, n=518  
M, F | Mean 72-78 | PA: a self-report on 4-point scale: very active, fairly active, not very active, not at all active) | 2.4 |
<table>
<thead>
<tr>
<th>Assessment of dementia</th>
<th>Covariates</th>
<th>Main results (ORs and 95% confidence intervals for dementia, low PA used as a reference group if not mentioned otherwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive tests: 1) Speed of processing: digit symbol substitution test, Figure Comparison, modified Stroop test part I and II, 2) memory: modified version of the California Verbal Learning Test, immediate and delayed recall, 3) executive function: Digits backwards, a shortened version of the CANTAB Spatial Working Memory test, the Stroop test part III, 4) MMSE and DSM-IV diagnosis of dementia</td>
<td>Age, sex, education, body mass index, systolic blood pressure, smoking, cholesterol</td>
<td>Medium PA: OR 0.59 (0.40–0.88) high PA: OR 0.76 (0.34–1.63)</td>
</tr>
<tr>
<td>Screening: MMSE and the Geriatric Mental Schedule (GMS) organic level. Diagnoses: the Cambridge Examination for Mental Disorders in the Elderly (CAMDEX) interview, DSM-III-R, NINCDS-ADRDA.</td>
<td>Age, sex, score on MMSE (not specified if baseline), low educational level, smoking, APOE ε4 carrier status, hypertension, BMI, diabetes, total cholesterol, HDL cholesterol</td>
<td>HR for dementia: 0.93 (0.85–1.02)</td>
</tr>
<tr>
<td>Medical records, Cambridge Cognitive Score, comprehensive diagnostics for those who screened positive in Cambridge Cognitive Score</td>
<td>Age, social class</td>
<td>OR 0.41 95% CI 0.22, 0.77; P&lt;0.005</td>
</tr>
<tr>
<td>The Telephone Interview for Cognitive Status (TICS), animal naming, phonemic fluency (count of F, A, and S words produced), Hopkins Verbal Learning Test (HVLT), and the Brief Test of Attention. Average of z-scores of tests.</td>
<td>Age, sex, smoking, diabetes, and hypertension in 1978</td>
<td>Regular exercise: HR for dementia: 1.51 (95% CI 0.55, 4.20) (&lt;45 y (n=285) HR: 2.37 (95% CI 0.29, 19.09) MET-hours per day: HR: 1.36 (95% CI 0.91, 2.03) (&lt;45 y (n=285) HR: 1.65 (95% CI 0.91, 3.01)</td>
</tr>
<tr>
<td>The 10-Word Delayed Recall Test and the Clinician Home-based Interview to assess Function</td>
<td>Age, sex, education</td>
<td>OR for dementia: High: 1.00, Moderate 1.0 (0.4–2.7), Low: 1.5 (0.5–4.6)</td>
</tr>
<tr>
<td>Screening: 3MS. Diagnosis of vascular dementia: DSM-IV and NINDS-AIREN.</td>
<td>Age, region</td>
<td>Vascular dementia: OR for men: 1.24 (0.57–2.94) and for women: 0.46 (0.25–0.82)</td>
</tr>
<tr>
<td>DSM-III</td>
<td>Age, sex, education, baseline MMSE score, comorbidity, physical functioning and depressive symptoms.</td>
<td>0.61 (0.42–0.87) (0 vs. 1-12 physical activity component score)</td>
</tr>
<tr>
<td>Screening: MMSE. Diagnoses: standard criteria for dementia.</td>
<td>Age, sex, education, depression, vascular risk scores, APOE genotype</td>
<td>OR for dementia (per descending category of activity i.e. very active, fairly active, not very active, not at all active): 2.72 (1.61–4.60)</td>
</tr>
<tr>
<td>Source/cohort</td>
<td>Subjects at baseline (N), subjects at follow-up (n), Sex (M for men and F for women)</td>
<td>Age at baseline</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>(Kishimoto et al., 2016) (Japan)</td>
<td>N=837, n=803, M, F</td>
<td>74</td>
</tr>
<tr>
<td>(Larson et al., 2006) (USA)</td>
<td>N=5422, n=1895, n=1740, M, F</td>
<td>74</td>
</tr>
<tr>
<td>(Laurin et al., 2001) (Canada)</td>
<td>N=6434, n=4615, M, F</td>
<td>≥ 65</td>
</tr>
<tr>
<td>(Lee, A. T. et al., 2015) (Hongkong)</td>
<td>N=15589, M, F</td>
<td>≥ 65</td>
</tr>
<tr>
<td>(Li, G. et al., 1991) (China)</td>
<td>n=1090, M, F</td>
<td>≥ 60</td>
</tr>
<tr>
<td>(Lindsay et al., 2002) (Canada)</td>
<td>N=6434, n=4615, M, F</td>
<td>≥ 65</td>
</tr>
<tr>
<td>(Llamas-Velasco et al., 2015) (Spain)</td>
<td>N=6395, n=5278 (baseline), n=3105, M, F</td>
<td>74,3</td>
</tr>
<tr>
<td>(Luck et al., 2014) (Germany)</td>
<td>N=22701, n=10850, n=2492, M, F</td>
<td>81,1</td>
</tr>
<tr>
<td>Assessment of dementia</td>
<td>Covariates</td>
<td>Main results (ORs and 95% confidence intervals for dementia, low PA used as a reference group if not mentioned otherwise)</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DSM-III, NINCDS-ADRDA, NINDS-AIREN for VaD or autopsy-based diagnoses</td>
<td>Age, sex low education level, systolic blood pressure, antihypertensive agents, diabetes, total cholesterol, body mass index, electrocardiogram abnormalities, history of stroke at entry, smoking habits, alcohol consumption</td>
<td>HR for active group 0.78 (0.60–1.01)</td>
</tr>
<tr>
<td>Screening: CASI. Diagnosis: DSM-IV, NINCDS-ADRDA.</td>
<td>Presence of apolipoprotein E ε4 alleles, diabetes, hypertension, cerebrovascular disease, coronary heart disease, self-rated health, physical performance, depression, cognitive functioning</td>
<td>HR: 0.68 (0.48–0.96)</td>
</tr>
<tr>
<td>3MS for screening, 3-stage clinical evaluation for diagnosis, NINCDS-ADRDA-criteria</td>
<td>Age, education, regular smoking, regular alcohol use, use of NSAID, functional ability in basic and instrumental activities of daily living, self-rated health, the number of chronic health conditions (only age and education for cognitive decline)</td>
<td>Dementia: high: 0.63 (0.40 –0.98)</td>
</tr>
<tr>
<td>Clinical dementia according to ICD-10 or Clinical Dementia Rating (CDR) I-3</td>
<td>Age, gender, educational level, and physical and psychiatric comorbidities.</td>
<td>OR for dementia for aerobic 0.81 (0.68–0.95) and mind-body exercises 0.76 (0.63-0.92).</td>
</tr>
<tr>
<td>Screening: MMSE and the Crichton Royal Behavior Rating Scale. Diagnoses: modified DSM-III.</td>
<td>? (full text not reached)</td>
<td>Higher risk of developing dementia (results not shown in abstract)</td>
</tr>
<tr>
<td>Screening: 3MS. Diagnosis of Alzheimer’s disease: DSM-IV.</td>
<td>Age, sex, education</td>
<td>OR: 0.69 (0.50–0.96)</td>
</tr>
<tr>
<td>Screening: Spanish adaptation of MMSE and an adapted Spanish version of Pfeffer’s Functional Activities Questionnaire. Diagnoses: DSM-IV, NINCDS-ADRDA</td>
<td>Age, sex, education, alcohol consumption, stroke, hypertension, Charlson index (“health indicator”)</td>
<td>HR for dementia: Light PA: 0.53 (0.34–0.82), Moderate PA: 0.45 (0.27–0.76), High PA: 0.29 (0.16–0.52)</td>
</tr>
<tr>
<td>DSM-IV, NINCDS-ADRDA</td>
<td>Age, gender, education, alcohol consumption, smoking, MMSE score, mental activity and co-morbidity at follow-up I (diabetes mellitus, hypertension, cardiac arrhythmia, coronary heart disease, myocardial infarction, peripheral arterial obstructive disease, carotid artery stenosis &gt; 80%, transient ischaemic attack, stroke, hyperlipidaemia, hypercholesterolaemia, hyperthyroidism, hypothyroidism, traumatic brain injury and depression)</td>
<td>0.79 (0.70–0.90)</td>
</tr>
</tbody>
</table>
Appendices

<table>
<thead>
<tr>
<th>Source/cohorts</th>
<th>Subjects at baseline (N), subjects at follow-up (n), Sex (M for men and F for women)</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (mean in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(McCallum et al., 2007) (Australia)</td>
<td>n=2805 (73% attendance rate baseline) M, F</td>
<td>&gt; 60</td>
<td>Interview: walking (daily vs. rarely)</td>
<td>14</td>
</tr>
<tr>
<td>(Morgan et al., 2012) (UK)</td>
<td>N=2959, n=1225 M</td>
<td>56</td>
<td>Minnesota Leisure Time Physical Activity Questionnaire: 1) high 2) moderate 3) low tertiles from a composite value combining activity type, duration and frequency</td>
<td>16</td>
</tr>
<tr>
<td>(Paganini-Hill et al., 2016) (USA)</td>
<td>N=1919, n=587 M, F</td>
<td>93</td>
<td>Doing vigorous exercise (rarely/never vs. more frequently)</td>
<td>3</td>
</tr>
<tr>
<td>(Podewils et al., 2005) (USA)</td>
<td>N=5888, n=3375 M, F</td>
<td>74,8</td>
<td>Quartiles of leisure-time energy expenditure</td>
<td>5,4</td>
</tr>
<tr>
<td>(Ravaglia et al., 2008) (Italy)</td>
<td>N=1353, n=749 M, F</td>
<td>73,2</td>
<td>1) Tertiles of weekly energy expenditure 2) ACSM recommendation of PA: yes or no</td>
<td>3,9</td>
</tr>
<tr>
<td>(Rosness et al., 2014) (Norway)</td>
<td>N=31086, n=26941, n=26055 (fully adjusted) M, F</td>
<td>72,4</td>
<td>Self-report of an average week fro the past year: Light exercise (not causing perspiration or panting) less or more than 3 hours per week, Hard exercise (causing panting and perspiration) less or more than 3 hours a week</td>
<td>10,3</td>
</tr>
<tr>
<td>(Rovio et al., 2007) (Finland)</td>
<td>N=2000, n=1158 M, F</td>
<td>50,4</td>
<td>Work-related PA: “How physically heavy is your work?” 1) sedentary work 2) physical work . Commuting PA to and from work: 1) not at all 2) 59min or less 3) at least 60 minutes</td>
<td>20,9</td>
</tr>
<tr>
<td>Assessment of dementia</td>
<td>Covariates</td>
<td>Main results (ORs and 95% confidence intervals for dementia, low PA used as a reference group if not mentioned otherwise)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical records: ICD-9, ICD-10</td>
<td>Marital status, education, prior history of stroke, activities of daily living</td>
<td>1.00 (0.78-1.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: CAMCOG, diagnosis: NINDS-AIREN criteria</td>
<td>Age, social class, National Adult Reading Test score, smoking status, marital status, self-reported history of vascular disease, alcohol consumption, body mass index, common mental disorder, Spielberger’s State-Trait Anxiety Index score</td>
<td>OR: moderate: 0.57 (0.28–1.16), high: 1.16 (0.61–2.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-IV for most (for whom in-person evaluation was possible), for others: MMSE (age-specific and education-specific cut-off score derived from the same cohort) or CASI ≤ 25</td>
<td>Age, sex, education</td>
<td>HR for incident dementia: rarely/never: 1.00, more frequently: 0.88 (0.67–1.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: the modified Mini-Mental State Examination (3MS)</td>
<td>Age, educational level, gender, ethnicity, apolipoprotein E genotype, baseline 3MS, baseline MRI white-matter-grade score, activities of daily living impairment, instrumental activities of daily living impairment, Lubben Social Network Score, social support score.</td>
<td>HR: lowest quartile: 1.00, second quartile: 0.94 (0.69–1.28), highest quartile: 0.85 (0.61–1.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: MMSE, ADL (activities of daily living); IADL (instrumental activities of daily living) Diagnoses: DSM-IV, NINCDS-ADRDA, NINDS-AIREN</td>
<td>Age, gender, education, APOE genotype, cardiovascular disease, hypertension, hyperhomocysteinemia, basic activities of daily living motor disability</td>
<td>HR for lowest tertile: 1.00, second tertile : 0.69 (0.41–1.15), highest tertile: 0.58 (0.32–1.06). ACSM recommendations: no: 1.00, yes: 0.73 (0.46–1.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norwegian Cause of Death Registry (ICD-10, F00-F03, G30.0-G30.9)</td>
<td>Age, gender, hypertension, total cholesterol, BMI, self-reported diabetes, smoking, education</td>
<td>HR: inactive: 1.00, light &lt; 3h/wk: 0.74 (0.62–0.88), Light &gt; 3h/wk: 0.61 (0.51–0.73), Hard &lt; 3 h/wk: 0.50 (0.41–0.61), Hard &gt; 3h/wk: 0.56 (0.43–0.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: MMSE, diagnoses according to the DSM-IV criteria and NINCDS-ADRDA criteria</td>
<td>Age, sex, education, follow-up time, locomotor symptoms, main occupation during life, income, leisure-time and commuting PA (or occupational PA), APOE ε4 genotype, body mass index, blood pressure, cholesterol, history of myocardial infarction, stroke and diabetes mellitus, smoking status</td>
<td>Occupational PA: OR 1.45 (0.66–3.17) Commuting PA: sedentary 0.58 (0.26–1.28), moderate: 1, active 0.46 (0.10–2.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source/cohort</td>
<td>Subjects at baseline (N), subjects at follow-up (n), Sex (M for men and F for women)</td>
<td>Age at baseline</td>
<td>Assessment of PA</td>
<td>Follow-up (mean in years)</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>(Rovio et al., 2005) (Sweden)</td>
<td>N= 2000, n=1449 M, F</td>
<td>50.6</td>
<td>Questionnaire: 1) PA causing breathlessness and sweating at least 2x/week 2) PA causing breathlessness and sweating less than 2x/week</td>
<td>21</td>
</tr>
<tr>
<td>(Scarmeas et al., 2009) (USA)</td>
<td>N=4165, n=1880 M, F</td>
<td>77.2</td>
<td>1) No PA: median of 0 hours per week, 2) Some PA: 0.1 h/week of vigorous, 0.8 h/week of moderate or 1.3h/week of light or a combination thereof, 3) Much PA: median of 1.3 h/week of vigorous, 2.4 h/week of moderate, 3.8 h/week of light or a combination thereof.</td>
<td>5.4</td>
</tr>
<tr>
<td>(Taaffe et al., 2008) (Hawaii, USA)</td>
<td>N=3734, n=2263 M</td>
<td>Mean age 76–79</td>
<td>Self-report of PA: 1) Low: ≤ 28.7 (oxygen consumption index per 24-hours) 2) Moderate: 28.8 -32.4 3) High: ≥ 32.5</td>
<td>6.1</td>
</tr>
<tr>
<td>(Tan et al., 2017) (USA)</td>
<td>n=3714 M, F</td>
<td>71</td>
<td>The PAI was a composite score constructed for each participant by weighting each hour in their typical day based on their activity level (based on oxygen consumption or metabolic equivalents) and summing up these weighted hours over a 24-hour period. Participants were asked to report the number of hours in a typical day spent sleeping (weighting factor (WF) = 1) and in sedentary (WF = 1.1), slight (WF = 1.5), moderate (WF = 2.4), and heavy activities (WF = 5). PAI quintiles.</td>
<td>7.5</td>
</tr>
<tr>
<td>Assessment of dementia</td>
<td>Covariates</td>
<td>Main results (ORs and 95% confidence intervals for dementia, low PA used as a reference group if not mentioned otherwise)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE for screening, thorough neurological and neuropsychological examinations, DSM-IV criteria</td>
<td>Age, sex, education, follow-up time, locomotor disorders, APOE genotype, vascular disorders (BMI, total serum cholesterol, systolic blood pressure, history of vascular myocardial infarction, stroke, diabetes), smoking (ever vs never), alcohol drinking (yes vs no)</td>
<td>Dementia: OR 0.47 (0.25–0.90) Alzheimer: OR 0.35 (0.16–0.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: 15 neuropsychological tests, CDR. Diagnoses: DSM-III (Revised), NINDS-ADRDA.</td>
<td>Cohort, age, sex, ethnicity, education, apolipoprotein E ε4 allele, caloric intake, BMI, smoking, depression, leisure activities (cognitive or social), comorbidity index, baseline CRD score, time between first dietary and first PA assessment.</td>
<td>HR: No PA: 1.00, Some PA: 0.71 (0.51–0.98), Much PA: 0.63 (0.45–0.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: CASI. Diagnoses DSM-III-R, NINCDS-ADRDA. the criteria for California Alzheimer’s Disease Diagnostic and Treatment Centers for vascular dementia.</td>
<td>Age, education, baseline CASI, BMI, midlife systolic and diastolic blood pressure, smoking status, cholesterol, hypertension, diabetes, coronary heart disease, depression.</td>
<td>HR for men with low physical function: low: 1.00, moderate: 0.43 (0.17–1.09), high: 0.57 (0.23–1.42), HR for ment with moderate physical function: low: 1.00, moderate: 1.03 (0.40–2.63), high: 0.49 (0.16–1.48), HR for ment with high physical function: low: 1.00, moderate: 0.56 (0.17–1.86), high: 1.57 (0.61–4.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dementia: DSM-IV, Alzheimer’s disease: NINCDS-ADRDA criteria</td>
<td>Age, sex, high school degree, APOEε4 allele status, log plasma homocysteine, systolic blood pressure, diastolic blood pressure, antihypertensive medication, total cholesterol, current smoking, prevalent cardiovascular disease, diabetes, stroke, and atrial fibrillation</td>
<td>HR for dementia (quintile) vs. quintiles 2–5, while quintile 1 is the lowest quintile: 1.47 (1.06–2.04)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Appendices**

<table>
<thead>
<tr>
<th>Source/cohort</th>
<th>Subjects at baseline (N), subjects at follow-up (n), Sex (M for men and F for women)</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (mean in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Tolppanen et al., 2015) (Finland)</td>
<td>N=3559, n=1432 M, F</td>
<td>51</td>
<td>Question: “How often LTPA lasting at least 20–30 minutes and causing breathlessness and sweating?” 1) high: at least 2–3 times a week 2) moderate: once a week or 2–3 times a month 3) low: few times a year or never</td>
<td>28</td>
</tr>
<tr>
<td>(Wang, H. X. et al., 2002) (Sweden)</td>
<td>N= 1375, n=776 M, F</td>
<td>81,1</td>
<td>PA (swimming, walking or gymnastics): frequency 1) none 2) less than daily 3) daily</td>
<td>6,4</td>
</tr>
<tr>
<td>(Wang, S. et al., 2014) (USA)</td>
<td>n=1299, n=1249 F</td>
<td>83,3</td>
<td>Self-reported blocks walked per day and analyzed by tertile</td>
<td>5</td>
</tr>
<tr>
<td>(Verghese et al., 2003) (USA)</td>
<td>N=488, n=469 M, F</td>
<td>Mean 79–80</td>
<td>Physical activity score: Ten physical activities (tennis, golf, swimming, bicycling, dancing, group exercises, team games, walking, climbing more than two flights of stairs, babysitting). For each activity: 7 points for daily participation, 4 points for participating several days a week, 1 point for weekly participation, 0 points for participating occasionally or never.</td>
<td>6,6</td>
</tr>
<tr>
<td>(Wilson et al., 2002) (USA)</td>
<td>N=3838, n= 835, M, F</td>
<td>76,0</td>
<td>Question: Total weekly hours of PA</td>
<td>4,1</td>
</tr>
<tr>
<td>Assessment of dementia</td>
<td>Covariates</td>
<td>Main results (ORs and 95% confidence intervals for dementia, low PA used as a reference group if not mentioned otherwise)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the re-examination: screening: MMSE, diagnosis: according to NINCDS-ADRDA criteria. If not in the re-examination: from registers</td>
<td>For the entire cohort: age, sex, education, midlife BMI, marital status, occupational PA level, smoking, cardiorespiratory and musculoskeletal conditions. For the re-examined: age, sex, education, midlife occupational and LTPA level, marital status, smoking, cardiorespiratory and musculoskeletal conditions, and change in BMI from mid- to late life.</td>
<td>HRs for the entire cohort: HRs with high PA level used as a reference group: moderate: 1.46 (1.08–1.99), low: 1.39 (0.99–1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-III-R</td>
<td>Age, sex, education, baseline MMSE, comorbidity, depressive symptoms, physical functioning</td>
<td>RR for dementia: No: 1, Less than daily: 0.97 (0.42–2.22), Daily: 0.41 (0.13–1.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening: 3MS, a measure of global cognition, digit span backwards, a test of working memory or attention, the California Verbal Learning Test II Short Form including a delayed recall portion, verbal fluency test (as many words as possible starting with the letter “f” in a minute, category fluency with naming as many vegetables as possible within a minute and Trails B. Diagnoses: DSM-IV (for dementia) and Petersen criteria (for MCI).</td>
<td>Age, education, baseline cognitive scores</td>
<td>OR for dementia: lowest tertile: reference, second tertile: 0.65 (0.44–0.97), highest tertile: 0.57 (0.38–0.85).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-III-R</td>
<td>Age, sex, educational level, the presence or absence of chronic medical illnesses, baseline score on the Blessed-Information-Memory-Concentration test.</td>
<td>Physical Activity Score (1-point increment): HR 1.00 (0.98–1.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINCDS-ADRDA</td>
<td>demographic variables (not specified), ≤4</td>
<td>OR for AD: 1.04 (0.98–1.10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendices

<table>
<thead>
<tr>
<th>Source/cohort</th>
<th>Subjects at baseline (N), subjects at follow-up (n), Sex (M for men and F for women)</th>
<th>Age at baseline</th>
<th>Assessment of PA</th>
<th>Follow-up (mean in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Yamada et al., 2003) (Japan)</td>
<td>N=2463, n=1774 M, F</td>
<td>30–79</td>
<td>Questionnaire: PA index calculated from occupational and leisure activities</td>
<td>25-30</td>
</tr>
<tr>
<td>(Yoshitake et al., 1995) (Japan)</td>
<td>N=887, n=885 M, F</td>
<td>Mean age 73–74</td>
<td>1) Physically active: daily exercise during the leisure period or moderate to severe PA at work 2) Not physically active</td>
<td>7</td>
</tr>
</tbody>
</table>

Abbreviations: ADL = Activities of daily living, CAMDEX = Cambridge Examination for Mental Disorders in the Elderly, CANTAB = Cambridge Neuropsychological Automated Battery, CASI = Cognitive Ability Screening Instrument, CDR = Clinical Dementia Rating, DSM-III-R = The Diagnostic and Statistical Manual of Mental Disorders 3rd edition revised, GMS = Geriatric Mental Schedule, HVLT = Hopkins Learning Verbal Test, IADL = Instrumental activities of daily living, IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly, NINDS-AIREN = National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherché et l'Enseignement en Neurosciences, TICS = modified TICS
<table>
<thead>
<tr>
<th>Assessment of dementia</th>
<th>Covariates</th>
<th>Main results (ORs and 95% confidence intervals for dementia, low PA used as a reference group if not mentioned otherwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening: CASI. Diagnoses: DSM-IV.</td>
<td>Age, sex, education</td>
<td>Results for PA not shown, but is said “no significant effect”</td>
</tr>
<tr>
<td>DSM-III-R, NINCDS-ADRDA or autopsy.</td>
<td>Age, sex, systolic blood pressure, stroke, alcohol consumption, Hasegawa’s dementia scale at baseline, diabetes, hematocrit</td>
<td>For VD: RR 0.81 (0.42–1.57) in age-adjusted model, for AD: RR 0.20 (0.06–0.68)</td>
</tr>
</tbody>
</table>
### Appendix III. Case-control studies assessing physical activity and cognition or dementia

<table>
<thead>
<tr>
<th>Source/cohort</th>
<th>Country</th>
<th>Subjects at baseline (N), subjects at follow-up (n)</th>
<th>Sex</th>
<th>Age at baseline</th>
<th>Assessment of physical activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Andel et al., 2008)</td>
<td>Sweden</td>
<td>N=4506, n=3134</td>
<td>M, F</td>
<td>48,1</td>
<td>1) hardly any 2) light exercise (such as walking or light gardening) 3) regular exercise involving sports 4) hard physical training</td>
</tr>
<tr>
<td>(Gelber et al., 2012)</td>
<td>Hawaii, USA</td>
<td>N=3734, n=3468</td>
<td>M</td>
<td>52</td>
<td>Structured interview: Highest quartile of slight or moderate PA, corresponding to a mean (SD) of 7.2 (3.2) hours typically spent in slight activity per day or 4.4 (3.0) hours in moderate activity</td>
</tr>
<tr>
<td>Follow-up (mean in years)</td>
<td>Assessment of cognitive function</td>
<td>Covariates</td>
<td>Main results (ORs and 95% confidence intervals for dementia, low physical activity used as a reference group if not mentioned otherwise)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>screening: telephone screening, diagnosis: in-person clinical evaluation</td>
<td>age at cognitive screening, gender, education, smoking, alcohol consumption, portion of fruits and vegetables in diet, BMI, angina pectoris</td>
<td>OR: hardly any: 1.00, light 0.63 (0.43–0.91), regular 0.34 (0.16–0.72), hard: 0.70 (0.43–1.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>DSM-III</td>
<td>age, years of education, APOE e4 status, childhood years spent in Japan, occupational status, high cholesterol, and history of hypertension, diabetes, and cardiovascular disease</td>
<td>For low PA: HR 1.59 (95% CI 1.15–2.18)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ORIGINAL ARTICLES

   – Reprinted with permission from Informa UK Limited, trading as Taylor & Francis Group, the article is available online at: https://doi.org/10.3109/07853890.2014.994675 or through http://www.tandfonline.com. Copyright received 2019.
