Review article

Twin studies on the association of physical activity with cognitive and cerebral outcomes

Paula Iso-Markku, Katja Waller, Pekka Hautasaari, Jaakko Kaprio, Urho M. Kujala, Ina M. Tarkka

A R T I C L E   I N F O
Keywords: Cognition, Age-related brain changes, Leisure-time physical activity, Structural MRI, Electroencephalography, Mismatch negativity, Somatosensory system

A B S T R A C T
Regular physical activity (PA) offers positive effects on the human body. However, the effects of PA on cognition and in the brain are less clear. In this paper, we narratively review the relationship of PA with cognition and dementia, first from general perspective and then through genetically informed studies on the topic. Then we move on to imaging studies on exercise and brain anatomy first by presenting an overall picture of the topic and then discussing brain imaging studies addressing PA and brain structure in twins in more detailed way. Regarding PA and cognition or dementia, genetically informed studies are uncommon, even though the relationship between PA and cognitive ageing has been extensively studied. It is challenging to find twin pairs discordant for PA and dementia. Concerning brain imaging studies, among PA discordant young adult twin pairs, the more active co-twins showed larger gray matter volumes in striatal, prefrontal, and hippocampal regions and in electrophysiological studies automatic deviance-detection processes differed in brain regions involved with sensorimotor, visual and memory functions.

1. Introduction

Physical activity (PA) is commonly defined as bodily movement produced by skeletal muscles that results in energy expenditure. Furthermore, PA in daily life may include occupational, sports, household and/or other activities. Exercise, on the other hand as a subset of physical activity, is planned, structured, and repetitive aiming to improve or maintain physical fitness (Caspersen et al., 1985). PA at large has several positive effects on the human body. Pulmonary function, cardiovascular system, skeletal muscles, and endocrine system of the body benefit from PA, it improves overall quality of life and reduces risk for obesity and is associated with low risk for several chronic diseases (McArdle et al., 2001). Leisure-time physical activity (LTPA), behaviors connected with PA performed in free time, is distinct from such PA which is part of gainful employment. Observational follow-up studies have shown associations between high LTPA and declined future occurrence of chronic diseases. Designs to provide stronger evidence for causality such as randomized controlled trials and co-twin control studies in monozygotic twins (MZ) have confirmed the effect of PA on functional outcomes and disease risk factors. In contrast MZ twin pair co-twin control designs have not been able to provide evidence for a causal relation between PA and morbidity or mortality (Leskinen and Kujala, 2015; Kujala, 2018) except for type 2 diabetes (Waller et al., 2010).

As a regular life-long habit PA is associated with individual’s childhood environment and childhood family behavioral patterns (Tammelin et al., 2003; Huppertz et al., 2017). The contribution of genetic factors to variations in PA can be studied with the help of twin studies, by comparing the relative similarity of MZ and dizygotic (DZ) pairs. If MZ and DZ pairs are equally similar for a behavior such as PA, it indicates that familial but not genetic factors, i.e. environmental influences shared by the twins (such as family socio-economic status, place of residence, etc.) are influencing the behavior (Boomsma et al., 2002; Posthuma et al., 2003; Kaprio and Silventoinen, 2011; van Dongen et al., 2012). On the other hand, greater similarity of MZ pairs compared to DZ pairs provides evidence for the role of genetic factors (Kujala et al., 2002; Zwijnenburg et al., 2010). The proportion of variance, i.e. interindividual differences in a trait such as PA that is
### Table 1

Randomized controlled trials of aerobic exercise and cognition.

<table>
<thead>
<tr>
<th>Meta-analyses on PA and cognition</th>
<th>Number of studies included</th>
<th>Mean sample size in the studies included</th>
<th>Type of participants</th>
<th>Mean intervention length in the studies included</th>
<th>Intervention</th>
<th>Effect size (and 95% confidence intervals)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith et al. (2010)</td>
<td>29</td>
<td>71</td>
<td>adults at least 18 years old, non-demented</td>
<td>4.5 months</td>
<td>Heterogenous</td>
<td>Eg. Hedge’s g for attention and processing speed 0.158 (0.055, 0.260)</td>
<td>Aerobic training was associated with modest improvements in attention and processing speed, executive function and memory but not with working memory.</td>
</tr>
<tr>
<td>Hindin and Zelinski (2012)</td>
<td>42</td>
<td>90</td>
<td>Healthy cognitively unimpaired community-dwelling age 55 + adults</td>
<td>3.3 months</td>
<td>Heterogenous</td>
<td>Between-group effect size 0.33 (0.10, 0.55)</td>
<td>Aerobic training improves performance on cognitive tasks.</td>
</tr>
<tr>
<td>Kelly et al. (2014)</td>
<td>19</td>
<td>78</td>
<td>Community-dwelling &gt; 50 years aged adults with no cognitive impairment</td>
<td>6.7 months</td>
<td>Heterogenous</td>
<td>Eg. standard mean difference for processing speed 0.27 (-0.22, 0.75)</td>
<td>No significant differences in exercise and controls on any of the comparison.</td>
</tr>
<tr>
<td>Young et al. (2015)</td>
<td>10</td>
<td>69</td>
<td>55 years old or older persons without objective cognitive impairment</td>
<td>3.9 months</td>
<td>Heterogenous</td>
<td>Eg. standard mean difference for cognitive speed in comparison to any active intervention: 0.12 (-0.08, 0.33)</td>
<td>No evidence for benefit of aerobic training in any of the cognitive domains studied (cognitive speed, verbal memory functions, visual memory functions, working memory, memory functions (delayed), executive function, perception, cognitive inhibition, visual attention, auditory attention, motor function).</td>
</tr>
</tbody>
</table>

**Note:** All studies were conducted with randomized controlled trials, except for the single largest randomized controlled trial on PA and cognition to our knowledge.

<table>
<thead>
<tr>
<th>Single largest randomized controlled trial on PA and cognition to our knowledge</th>
<th>Sample size</th>
<th>Type of participants</th>
<th>Intervention length</th>
<th>Intervention</th>
<th>Effect size (and 95% confidence intervals)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sink et al. (2015)</td>
<td>818</td>
<td>Sedentary adults 70 – 89 years old who were at risk of mobility disability but were able to walk 400 m</td>
<td>24 months</td>
<td>A structured moderate-intensity program that included walking, resistance training and flexibility exercises</td>
<td>Eg. mean difference between exercise and health education groups in Digit Symbol Test -0.01 (-0.80, 0.77)</td>
<td>Physical activity program showed no improvements in global cognition or domain-specific cognition.</td>
</tr>
</tbody>
</table>
accounted for by interindividual differences in genetics is known as the heritability of a trait. Heritability estimates are population and time dependent and are not constant properties of the trait. Heritability can also change during different periods of development and ageing. It is often observed that childhood environment plays an important role, but the contribution of shared environments is not long-lasting, such that genetic factors and environments unique to the individual (and by definition not shared with a sibling or family members) become more important with age. In Finnish data, Aaltonen et al. (2013) showed that MZ twins were more likely to have a similar leisure-time physical activity, LTPA, level than DZ twins. For twins, shared environmental influences on LTPA were also relatively stable during adolescence and these increased in young adulthood and according to Aaltonen et al. (2013), the heritability estimates of LTPA ranged between 43 % and 52 % among adolescents and declined to 30 % among young adults. Thus, when studying the association of PA with anatomical, physiological or cognitive indicators of the brain, teasing out the causal contribution of PA is challenging as the association may be confounded by genetic factors.

For further studies of anatomic, physiologic and cognitive effects of PA the co-twin study design is productive as it controls the contribution from genetic factors; MZ twins have the same genetic make-up and so any difference between them is presumably associated with PA and related factors that correlate with PA independent from sequence variation in the genome. However, finding MZ twin pairs discordant for PA for a relatively long time is a challenging task as PA varies over time, and twin pairs can be similar due to both genetic factors and shared environments. An example of the latter may be having the same or related occupation. On the other hand, discordance may arise for reasons unrelated to PA directly, such as a chronic disease or injury. Thus, we see more discordance for PA with age, but the consequences and brain effects of such PA discordance may be different from discordance that has arisen gradually over time with no obvious trigger.

In the present paper, we will narratively review recent findings of the relationship of PA with cognition and dementia in general and through genetically informed studies. First, we will discuss more broadly the issue of PA and cognitive decline in ageing and connect this issue with twin studies which have looked at the relationship of PA and cognitive ageing. Second, we will introduce the subject of imaging studies on PA and brain structure and discuss small exploratory studies of brain structural data in young twins using gray matter (GM) volume as an outcome measure. Finally, we draw attention to twin studies with sensitive brain electrophysiological measure, utilizing mismatch response as an outcome measure. We use the term ‘exercise’ when speaking about randomized controlled interventions, as there exercise is tailored for a specific purpose. When speaking about meta-analyses of observational studies we speak about PA as the recording of modes of PA vary in the studies included in the reviews. When speaking about specific studies we use LTPA, when and only when the questionnaire-based PA measure clearly refers to LTPA.

2. Associations of PA with cognition and ageing

Potential of PA to prevent dementia has been much debated in the neuroscience during the past years. Rodent studies have provided promising results in showing increased neurogenesis in the hippocampus (van Praag et al., 1999), increased vasculature (Lopez-Lopez et al., 2004), synaptic plasticity (Huttenrauch et al., 2016) and even decreased amyloid plaque formation (Adlard et al., 2005; Lazarov et al., 2005). While increased vasculature has also been shown to occur in response to exercise in humans (Bullitt et al., 2009), debate on human adult neurogenesis continues (Spalding et al., 2013; Sorrells et al., 2018) and clear evidence on the association of exercise with decreased amyloid deposition in humans lacks evidence though positron emission tomography imaging provides a new in vivo means to study amyloid depositions. The results in studies addressing amyloid accumulation in humans and PA have been conflicting (lower amyloid burden in two studies (Liang et al., 2010; Müller et al., 2018), not associated with in two studies (Vemuri et al., 2016; Souto Barreto et al., 2015), decreased amyloid burden only in a subgroup (Brown et al., 2013; Head et al., 2012; Brown et al., 2017). While much enthusiasm is focused on the topic, the evidence is not that clear in humans. Here we review observational studies and randomized controlled trials on PA and cognition or dementia in humans. We have not performed a systematic review and the studies presented here are discussed in a narrative manner.

2.1. Clinical trials

Prospective studies can do only so much to elicit indication of causality. Randomized controlled trials are the golden standard in attempts to detect causality. The randomized controlled trials addressing exercise and cognition with long interventions and large sample sizes are few. Most randomized controlled trials addressing solely exercise and cognition have been quite small in sample size and the intervention length has been quite modest when considering the long preclinical process of dementia (Table 1). The mean sample size in the studies of the meta-analyses addressing this issue has varied between 71–90 participants and the mean intervention length between 3.3 and 6.7 months (Young et al., 2015; Kelly et al., 2014; Hindin and Zelinski, 2012; Smith et al., 2010). Although the combined number of participants is considerably large, the meta-analysis is only as solid as the quality of the trials included in it and the small sample sizes in these trials underscore the quality of these meta-analyses. Moreover, the results have been conflicting: (beneficial effect on cognition: Smith et al., 2010; Hindin and Zelinski, 2012, no significant effect on cognition: Young et al., 2015; Kelly et al., 2014). The single largest randomized controlled trial examined the effect of 2-year physical activity intervention combining aerobic training, resistance training, balance exercises and flexibility exercises on global cognition and different cognitive domains in a sample of 818 sedentary 70–89 year-old adults (Sink et al., 2015). Their results and the fact that there are only few large randomized controlled trials addressing exercise alone in a long follow-up underline a different message from the animal models: there is no solid evidence for causal relationship between exercise and dementia prevention in humans.

2.2. Prospective studies of cognitive decline

Interestingly, the meta-analyses addressing prospective studies examining the relationship between PA and cognition tell a uniform story: they all suggest almost 40 % risk reduction for cognitive decline with high level of PA (Sofi et al., 2011; Blondell et al., 2014; Morgan et al., 2012; Guerre et al., 2017) (see Table 2). Some methodological aspects, however, limit the level of evidence from these meta-analyses. Two of these meta-analyses assessed the quality of the studies in their meta-analyses (Guerre et al., 2017; Blondell et al., 2014) and two did not (Sofi et al., 2011; Morgan et al., 2012). Irrespective of the study quality, both meta-analyses assessing quality (Guerre et al., 2017; Blondell et al., 2014) combined all studies they found. Low-quality studies in these meta-analyses limit the level of evidence they provide. Because the preclinical phase of dementia most probably lasts for decades (Bateman et al., 2012; Fagan et al., 2014), it could be that an association between physical inactivity and cognitive decline found in an elderly cohort with a short follow-up reflects rather an on-going dementing process than a causal association. The majority of the studies addressing PA and cognition in these meta-analyses have been implemented in elderly populations (mean baseline age ≥65 years in 7/8, 16/22, 15/21 and 13/15 of the cohorts included in the meta-analysis from Morgan et al. (2012), Guerre et al. (2017), Blondell et al. (2014) and Sofi et al. (2011), respectively) with short follow-ups (mean follow-up < 10 years in 13/15, 16/21, 20/22 and 7/8 in the studies included in the meta-analysis from
<table>
<thead>
<tr>
<th>Meta-analyses on PA and dementia</th>
<th>Pooled RR (95% CI)</th>
<th>Number of studies included</th>
<th>Hetero-genity</th>
<th>Evidence for publication bias</th>
<th>Subgroup analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamer and Chida (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 (0.63–0.87)</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 (9 cohorts)</td>
<td>No</td>
<td>Not evaluated</td>
<td>Association significant only in studies &lt; 10 years of duration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meta-analyses on PA and cognition</th>
<th>Pooled RR (95% CI)</th>
<th>Number of studies included</th>
<th>Hetero-genity</th>
<th>Evidence for publication bias</th>
<th>Subgroup analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sofi et al. (2011)</td>
<td>0.62 (0.54–0.70)</td>
<td>15 (12 cohorts)</td>
<td>No</td>
<td>No</td>
<td>Smaller studies and studies with a longer follow-up (&gt; 5 years) showed higher estimates of association</td>
</tr>
<tr>
<td>Morgan et al. (2012)</td>
<td>0.66 (0.52–0.85)</td>
<td>9</td>
<td>Yes</td>
<td>No</td>
<td>Association significant only in studies &lt; 5 years of duration</td>
</tr>
<tr>
<td>Blondell et al. (2014)</td>
<td>0.65 (0.55–0.76)</td>
<td>17 (21 cohorts)</td>
<td>Moderate</td>
<td>No</td>
<td>Association only significant in studies &lt; 10 years of follow-up</td>
</tr>
<tr>
<td>Guure et al. (2017)</td>
<td>0.67 (0.55–0.78)</td>
<td>18 (22 cohorts)</td>
<td>No</td>
<td>No</td>
<td>Association only significant in studies with cohorts at least 65 years at baseline</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meta-analyses on PA and AD</th>
<th>Pooled RR (95% CI)</th>
<th>Number of studies included</th>
<th>Hetero-genity</th>
<th>Evidence for publication bias</th>
<th>Subgroup analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davgulsh et al. (2011)</td>
<td>0.72 (0.53–0.98)</td>
<td>9</td>
<td>Yes</td>
<td>Not told</td>
<td>No subgroup analysis</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>No subgroup analysis</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>No subgroup analysis</td>
</tr>
<tr>
<td>Santos-Lorano et al. (2016)</td>
<td>0.65 (0.56–0.74)</td>
<td>10</td>
<td>Moderate</td>
<td>No</td>
<td>A subgroup analysis addressing PA according to international PA recommendations: significant risk reduction for physically active according to the recommendations</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>Association significant also in follow-ups &gt; 5 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meta-analyses on PA and vascular dementia</th>
<th>Pooled RR (95% CI)</th>
<th>Number of studies included</th>
<th>Hetero-genity</th>
<th>Evidence for publication bias</th>
<th>Subgroup analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarsland et al. (2010)</td>
<td>0.62 (0.42–0.92)</td>
<td>5</td>
<td>Yes</td>
<td>Some</td>
<td>No subgroup analysis</td>
</tr>
<tr>
<td>Guure et al. (2017)</td>
<td>0.92 (0.62–1.30)</td>
<td>8 (7 prospective studies)</td>
<td>Not significantly</td>
<td>Negligible</td>
<td>No subgroup analysis</td>
</tr>
</tbody>
</table>
Sofi et al. (2011), Blondell et al. (2014), Guure et al. (2017) and Morgan et al. (2012), respectively). Furthermore, sensitivity analyses show weaker or non-significant results in studies of better quality, longer follow-ups and cohorts in which baseline age is under 65 years in these analyses, it can be concluded that the likelihood of the results of these meta-analyses being biased by reverse causation is considerable as has been suggested also earlier (Morgan et al., 2012; Sabia et al., 2017; Kivimäki et al., 2019).

Prospective studies addressing midlife PA and late-life cognition with a long follow-up exceeding 10 years are scarce (see Supplementary material, Table 4). If we exclude studies that do not have sound measures of PA and cognition or do not control for the most important confounding factors in some fashion (age, sex and education or premorbid intelligence), the number is even smaller. By rigorous measures, we mean either objectively measured PA or a valid structured PA questionnaire from which it is possible to estimate the amount of PA in contrast to simple “yes” or “no” questions and to questionnaires assessing the diversity of PA. In the studies from Singh-Manoux et al. (2005), Chang et al. (2010) and Elwood et al. (2013) the average baseline age of the cohort ranged between 36 and 57 years, the length of the follow-up was between 11 and 30 years, the measures of PA and cognition were rigorous, and they controlled for the effect of age, sex, and education (Chang et al., 2010) or premorbid intelligence (Elwood et al., 2013; Singh-Manoux et al., 2005). All of these studies found a statistically significant association between midlife PA and late-life cognition. In the study by Singh-Manoux et al. (2005), high level of midlife PA was associated with fluid intelligence but not with memory and phonemic or semantic fluency. In the study by Chang et al. (2010), high level of midlife PA was associated with processing speed, executive function and memory. In the study from Elwood et al. (2013), midlife regular exercisers had a lower risk of developing cognitive decline. The effect of education on the risk of developing dementia is substantial (Meng and D’Arcy, 2012; Vuoksimaa et al., 2016) and most often controlled for in studies addressing other protective factors of cognitive decline. However, the studies from Elwood et al. (2013) and from Singh-Manoux et al. (2005) are exceptional because they also controlled for premorbid intelligence levels. Most studies addressing midlife PA and cognitive decline have not been able to consider midlife general cognitive ability (see Supplementary material, Table 4). This is a limitation, because in adolescents it has been shown that higher academic performance predicts participation in LTPA later in life but not vice versa (Aaltoinen et al., 2016). Additionally, in an elderly Finnish cohort executive function predicted mobility better than mobility predicted executive function in a 2-year follow-up (Poranen-Clark et al., 2018). If cognition predicts PA levels rather than vice versa in adolescents and in the elderly, then there is a good possibility that general cognitive ability predicts PA levels also from midlife into late-life. Most studies addressing midlife PA and late-life cognition do not take premorbid intelligence into account (see Supplementary material, Table 4), the two studies presented (Elwood et al., 2013; Singh-Manoux et al., 2005) here that did, found a significant association of midlife PA and late-life cognition. An interesting question is if this could be replicated in other cohorts.

Studies addressing objectively measured PA and cognition are so far scarce. To our knowledge, four longitudinal studies deal with elderly cohorts at least 65 years old with short 3–5 y follow-ups. Middleton et al. (2011) and Buchman et al. (2012) both found a significant inverse association between total PA and cognitive decline while Zhu et al. (2017) and Halloway et al. (2017) both found that some aspect of objectively measured PA was significantly associated with later better cognition. All cross-sectional studies addressing PA and cognition found a statistically significant association between some aspect of objectively measured PA and cognition or some aspect of cognition so far without clear consistency (Barnes et al., 2008; Zhu et al., 2015; Kerr et al., 2013; Johnson et al., 2016; Wilbur et al., 2012).

2.3. Prospective studies of dementia incidence

The high-quality (follow-up over 10 years, no significant selection bias in the study cohort, a thorough measure of PA, the result controlled for at least age, sex, education or a measure of midlife general cognitive ability and vascular morbidity) prospective studies addressing midlife PA and dementia show inconsistent results (see Supplementary material, Tables 5–6) Gelber et al. (2012) performed a case-control study of over 3400 Japanese-American men with a 25-year-long follow-up showing that PA among other healthy lifestyle choices is associated with lower risk for dementia. Andel et al. (2008) conducted a large cohort study of over 3000 twins with an extensive 31-year-long follow-up and showed a significant dementia risk reduction with PA at individual level which was not replicated in the co-twin control analyses. Rovio et al. (2010) studied the effect of LTPA that causes breathlessness and sweating on dementia incidence in a population-based sample of 1449 Finnish adults in a long follow-up exceeding 20 years. They showed that midlife LTPA was associated with decreased dementia incidence. Morgan et al. (2012) studied the effect of LTPA on dementia incidence in a population-based sample of 1225 adults in a 16-year-long follow-up finding no significant effect on dementia incidence. Chang et al. (2010) studied the amount of exercise on several different cognitive domains in 4945 Icelanders with an average of 26 years follow-up. They found that moderate amount of PA was associated with decreased amount of dementia incidence but high amount of PA (> 5 h/week) was not. Tolppanen et al. (2015) used the same cohort as Rovio et al. (2010) but focused also on the effect of late-life LTPA and the modifying effect of BMI. The association between increased dementia incidence and low level of midlife PA was only significant for men. Thus, conflict between the results on this topic prevails.

The meta-analyses addressing PA and dementia struggle with the same issues as meta-analyses from prospective studies addressing PA and cognitive decline. Majority of the studies have been implemented in elderly cohorts with short follow-ups endangering them for bias from reverse causality. The meta-analyses addressing PA and dementia all except one (Hammer and Chida, 2009) report either significant risk of publication bias (Morgan et al., 2012) or there was no significant inverse association if follow-ups were longer than 10 years (Blondell et al., 2011; Kivimäki et al., 2019) or if the baseline age of the participants was under 65 years (Guure et al., 2017; Xu et al., 2017) (see Table 2). However, in the recent meta-analysis by Xu et al. (2017), PA was nearly statistically significantly associated with decreased dementia incidence also in younger participants (< 65 years) and significantly associated with decreased dementia incidence in follow-ups over 10 years.

All meta-analyses addressing PA and Alzheimer’s Disease (AD) point to an inverse association (Beckett et al., 2015; Beydoun et al., 2014; Davglus et al., 2011; Santos-Lozano et al., 2016) (see Table 2). However, only two of these meta-analyses performed a quality assessment (Santos-Lozano et al., 2016; Guure et al., 2017) and both quality assessment tools left room for subjective evaluation not specifying clearly what was required for a reliable measure of PA, how long high-quality follow-up needed to be and what confounding factors were required to be taken into account for high-quality. For example, in the meta-analysis from Santos-Lozano et al. (2016) follow-up time was 5 years or longer in all but 2 studies, the quality was deemed to be high but the participants were on average from 70 to 80 year-olds. Meta-analyses addressing PA and vascular dementia suggest that there is no significant association (Aarsland et al., 2016; Guure et al., 2017) (see Table 2). These meta-analyses lean on quite a limited number of studies (5 studies in the meta-analysis from Aarsland et al. (2010) and 7 studies in the meta-analysis from Guure et al. (2017)). In the meta-analysis from Aarsland et al. (2010), the follow-up was 5.7 years, mean age at follow-up was over 70 years in all of the studies included and they were not able to take into account any confounding factors. In the meta-analysis from Guure et al. (2017), the mean follow-up was 9 years and all of the
studies included were based on elderly populations at least 65 years old. In addition to these meta-analyses, a plethora of reviews on PA and dementia exist. Many are narrative and do not employ systematic methods. Stephen et al. (2017) published a very robust systematic review with very specific quality assessment concluding that there is moderate quality evidence for an inverse association between PA and AD. Instead of regarding the studies implemented in old age being liable for reverse causation, they interpreted them as evidence for an extended window of opportunity for AD prevention with PA.

2.4. Twin studies addressing PA with cognition and ageing

As both PA and dementia have their own genetic properties, with modest to moderate heritability estimates (for dementia: Raiha et al., 1996; Gatz et al., 1997; Ferencz and Gerritsen, 2015 and for PA: Stubble et al., 2006; Lightfoot et al., 2018), there is potentially genetic confounding underlying the observed association between PA and dementia in observational studies. This genetic confounding may be due to multiple, polygenic effects shared by the genetic predisposition to exercise and the genetic predisposition to cognitive decline. In addition to genetic factors, other shared environmental factors may be influential. Kulshreshtha et al. (2019) showed in cross-sectional twin study design (544 MZ and DZ twin individuals) that familial factors other than genetics explain a large part of the association between a cardiovascular health score and cognition. This could also apply to the association of PA and cognitive ageing. Unfortunately, genetically informed studies of PA and cognition or dementia provided the possibility to address possible underlying genetics and other familial influences are few. The findings, however, offer a clear consistency (Table 3). While the results in the analyses of the whole cohort or case-control study vary over-20-year follow-ups, none of the twin studies has so far been able to show a significant association between PA and cognition or dementia in a co-twin control design controlling for genetic factors and shared environment (Andel et al., 2008; Carlson et al., 2008; Virta et al., 2015; Iso-Markku et al., 2016; Lee et al., 2014) except for Gatz et al. (2006), who were able to find a significant association between midlife PA and dementia other than AD type in a co-twin control design of 55 MZ twin pairs. However, the result was not significant for dementia in general or AD. Dementia other than AD type denoted largely, but not exclusively, vascular dementia in their study.

While the trust in the preventive power of PA has been strong, the results have been interpreted in favour of PA. Iso-Markku et al. (2015, 2016), Andel et al. (2008) and Virta et al. (2013) all hypothesize that the null results are due to low statistical power caused by the small number of twins discordant for both PA and dementia or cognitive impairment while Gatz et al. (2006) suspect that their minimal measure of PA (simple four-scale question) may detract from finding greater effects. The null results in co-twin control design might also reflect that the association between PA and cognition or dementia seen in other studies not taking genetics and shared environment into account is partly explained by them, at least for the most common cause of dementia: AD. The exceptional finding from Gatz et al. (2006), that PA was associated with decreased incidence of dementia other than AD type in a genetically controlled MZ co-twin control design, is promising in the battle against vascular dementia and suggests that the effect of PA on dementia may be type-specific. However, the finding needs to be replicated and it would be advisable to have a more exact measure of PA taking into account the amount and intensity of PA. To conclude, we have extensive amount of evidence showing a moderate strength association between PA and decreased dementia incidence in observational studies, especially in short follow-ups with elderly populations. When considering the association between midlife PA and late-life cognition and dementia, we have some evidence towards a cognition-protective relationship, but it is still weakened by the lack of controlling for genetic confounding. The evidence from genetically informed studies is even more infrequent and seems to support cognition-protective effect of PA only for vascular dementia. For other types of dementia, the positive association between PA and cognition seems to be at least partly explained by genetics and shared environmental factors. Interestingly, a recent GWAS (Visscher et al., 2017) showed that an ApoE gene variant associated with greater risk of AD is also associated with greater amount of moderate-to-vigorous PA (Klimentidis et al., 2018). This study is first to show that the genetic background of dementia is entwined to at least some extent with the genetic background of PA, even though the direction of this finding is somewhat surprising. One would expect that a variant in APOE ε4 allele, which is strongly implicated in AD, would be associated with sedentarity instead of greater amount of PA. In future studies, it is essential to focus on study quality, and to aim at revealing mutual temporal associations of both cognition and PA and to acquire more genetically informed high-quality studies on this issue.

3. Associations of PA with cortical structure and function

It is well accepted that exercise and functional training has all-important role in the recovery of stroke patients through brain plasticity (e.g. Hara, 2015). PA has also been shown to induce macro-scale anatomical changes in areas involved in motor-related functions (Lerch et al., 2011). It is not clear if exercise acts as a general brain plasticity booster in brain areas not contributing to movement in healthy adults. Furthermore, does PA prevent brain atrophy progression in old age? Cross-sectional studies provide reassuring results for both healthy adolescents (Chaddock et al., 2010) and for older adults (Colcombe et al., 2003). Causal interpretations are naturally impossible in these study designs. PA has been shown to be associated with total gray matter (GM) volume in longitudinal studies (Erickson et al., 2010; Ravio et al., 2010). The studies are limited, however, with the baseline level of the volumes being unknown leaving room for the possibility of reverse causation. A large study addressing objectively measured PA and frontal lobe progression in middle-aged and elderly cohort in 8-year follow-up found an association between amount of PA and decreased frontal lobe atrophy progression but not with temporal lobe atrophy progression (Yuki et al., 2012). However, this large longitudinal study did not control their results for baseline cerebral volumes although they did exclude severe brain atrophy cases from their study to reduce the possibility of reverse causation. Causal inference from randomized controlled trials is difficult because of conflicting results (Colcombe et al., 2006; Stephen et al., 2019) and different studies show volumetric increase in varying regions of gray matter without consistency (Batouli and Saba, 2017). At least two randomized controlled trials have not shown increase in white matter volume following exercise intervention (Colcombe et al., 2006) or an intervention including exercise (Stephen et al., 2019). Next, we are going to discuss the twin study results on this topic because they provide an outstanding view on the matter when controlling for genetic factors and shared environmental effects.

3.1. Brain structure and PA in twins

It is well agreed in magnetic resonance imaging (MRI) studies that total brain volume, and left and right hemisphere volumes present high heritability in adult twins (Lessov-Slaggar et al., 2012; Jansen et al., 2015). Earlier evidence from twin studies shows that also white matter (WM) volume is highly heritable (Carmelli et al., 1998), but hippocampal size is more affected by environmental influences (Sullivan et al., 2001). Only few studies have investigated the associations of twins, human brain imaging data and exercise. Few studies do, however, address these relationships. A co-twin control study from the United States performed a cross-sectional analysis of self-reported PA and brain morphology (n = 74 elderly MZ twins) (Carmelli et al., 1999). They also examined many midlife cardiovascular risk factors, but the data on current PA was at the time the MRI was performed.
### Twin studies of PA, cognition, and dementia.

<table>
<thead>
<tr>
<th>Longitudinal cohort studies</th>
<th>Sample size</th>
<th>Type of participants</th>
<th>Length of follow-up</th>
<th>Measure of physical activity</th>
<th>Measure of cognition</th>
<th>Effect size (95% confidence intervals)</th>
<th>Results from co-twin control design</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virta et al. (2013)</td>
<td>2165</td>
<td>Approximately 52-year-old Finnish twins</td>
<td>23 years</td>
<td>Detailed questionnaire on duration, frequency and intensity</td>
<td>TELE</td>
<td>Risk of cognitive impairment for sedentary (vs. conditioner)</td>
<td>2.52 (1.10 – 5.76)</td>
<td>Low leisure-time physical activity associated with decreased incidence of cognitive impairment, similar trend in within-pair analyses</td>
</tr>
<tr>
<td>Iso-Markku et al. (2015)</td>
<td>21,791</td>
<td>24 – 60-year-old Finnish twins</td>
<td>29 years</td>
<td>Detailed questionnaire on duration, frequency and intensity</td>
<td>TELE and TICS</td>
<td>Risk for dementia mortality for long-term vigorously active (vs. long-term inactivity)</td>
<td>0.65 (95% CI 0.43 – 0.98)</td>
<td>Midlife long-term vigorous PA was associated with decreased incidence of cognitive impairment, similar trend in within-pair analyses</td>
</tr>
<tr>
<td>Iso-Markku et al. (2016)</td>
<td>3,050</td>
<td>Finnish twins (mean age 49 years at baseline)</td>
<td>25 years</td>
<td>Detailed questionnaire on duration, frequency and intensity</td>
<td>TELE and TICS</td>
<td>Risk for cognitive impairment for long-term vigorous PA (vs. inactivity)</td>
<td>OR 0.50, 95% CI 0.35 – 0.73</td>
<td>Midlife long-term vigorous PA was associated with decreased incidence of cognitive impairment, similar trend in within-pair analyses</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case-control studies</th>
<th>Sample size</th>
<th>Type of participants</th>
<th>Length of follow-up</th>
<th>Measure of physical activity</th>
<th>Measure of cognition</th>
<th>Effect size (95% confidence intervals)</th>
<th>Results from co-twin control design</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gatz et al. (2006)</td>
<td>3,373</td>
<td>Middle-aged Swedish twins</td>
<td>30 years</td>
<td>Physical exercise: hardly any, light, regular, hard</td>
<td>Telephone screening followed by clinical diagnostic evaluation</td>
<td>Risk for total dementia low vs. high</td>
<td>1.00 (0.71, 1.42)</td>
<td>Physical exercise was significantly protective of dementia other than AD after controlling for genetic influences</td>
</tr>
<tr>
<td>Carlson et al. (2008)</td>
<td>1,47</td>
<td>Male twins from Registry of World War II veteran male twins (mean age 45 at baseline)</td>
<td>28 years</td>
<td>PA (max score 4) if participation in all four activities: &quot;outdoor activities&quot;, &quot;sports&quot;, &quot;gardening and home improvement&quot; and &quot;physical exercise after age 35&quot;</td>
<td>Multistep evaluation with clinical evaluation</td>
<td>Risk for dementia for higher PA (OR): 0.99</td>
<td>0.85 (0.56-1.29)</td>
<td>Exercise at midlife may reduce the risk of dementia other than AD</td>
</tr>
<tr>
<td>Andel et al. (2008)</td>
<td>3,134</td>
<td>Swedish twins (mean age 48 years at baseline)</td>
<td>31 years</td>
<td>Same as in Gatz et al. (2006)</td>
<td>Telephone screening and clinical evaluation</td>
<td>Risk for dementia (OR): 0.34</td>
<td>0.24-1.83, n = 84 MZ twins</td>
<td>Exercise at midlife is protective of dementia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cross-sectional studies</th>
<th>Sample size</th>
<th>Type of participants</th>
<th>Measure of physical activity</th>
<th>Measure of cognition</th>
<th>Effect size (95% confidence intervals)</th>
<th>Results from co-twin control design</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iso-Markku et al. (2015)</td>
<td>27,233</td>
<td>Community-dwelling Finnish twins (mean age 75)</td>
<td>TELE and TICS</td>
<td>Number of activities, the intensity of each activity, the frequency of participation</td>
<td>Risk for dementia (OR): 0.66</td>
<td>0.27-1.35, n = 64 MZ twins</td>
<td>Exercise at midlife is protective of dementia</td>
</tr>
<tr>
<td>Lar et al. (2014)</td>
<td>119</td>
<td>Community-dwelling Swedish twins (mean age 71)</td>
<td>TELE and TICS</td>
<td>Number of activities, the intensity of each activity, the frequency of participation</td>
<td>Risk for dementia (OR): 0.49</td>
<td>0.17-1.48, n = 119</td>
<td>Exercise at midlife is protective of dementia</td>
</tr>
</tbody>
</table>

**Abbreviations:** OR = odds ratio, TELE = Telephone Assessment for dementia, TICS = Telephone Interview for Cognitive Status, MZ = monozygotic.
Although many other cardiovascular risk factors from midlife were associated with greater amount of white matter hyperintensities and brain parenchyma volume independent of genetics and shared environmental influences, the current PA was not associated with total brain volume or the amount of WM hyperintensities. PA was, however, associated with the cerebrospinal fluid volume in within-pair analyses. The implication of this finding is less clear.

We have looked at MRIs of healthy MZ twin pairs (mean age 34.5 ± 1.5 y) who were discordant in their long-term PA, including LTPA (Rottensteiner et al., 2015; Hautasaari et al., 2017; Tarkka et al., 2019). Our sample of twin brothers was a subgroup from the population-based longitudinal study, FinnTwin16 Study, on Finnish twins born between October 1974 and December 1979 (Kaidesoja et al., 2019). This subgroup included a total of 22 twin pairs who underwent MRI and EEG measurements. The rationale detecting their discordance in PA has been described earlier in detail (Rottensteiner et al., 2015). Twin pairs were divided into more and less active co-twin based on multiple interviews and questionnaires over time regarding PA. All 22 pairs with MRI data were at least 1 year or longer PA discordant and were also different in their body composition and fitness level (significant difference in the total body fat% registered with dual-energy X-ray absorptiometry, DXA, and in maximal oxygen consumption, VO2max). The most common types of PA among the participants were jogging and walking.

Total-brain WM, GM and total intracranial volumes, as estimated from non-normalized images, did not differ between more and less active co-twins. Interestingly, there was regional differentiation in GM volumes (analyzed with whole-brain voxel-based morphometry) between more and less active co-twins suggesting higher GM volume in the left hippocampus in more active co-twins in a cohort of 22 MZ pairs (Tarkka et al., 2019). The sub-sample of 9 MZ twin pairs from the original cohort were discordant for PA, including LTPA, for over three years and among these 9 the active co-twin of the twin pair showed larger striatal GM volume and the non-dominant inferior frontal gyrus had also a larger GM volume compared to the inactive co-twin (Rottensteiner et al., 2015). As total cortical volumes are typically stable in early adulthood (Jansen et al., 2015), these findings provide evidence for structural modulation in healthy young adult brain possibly associated with long-term PA.

Utilizing heritability studies with MZ and DZ twin pairs may add to our understanding of individual structural brain differences and conceivable extent of adult neuroplasticity associated with PA. This would mean conducting bivariate genetic modelling to tease apart the phenotypic association between PA and structural brain differences in their genetic and environmental components and estimate the genetic correlation between the genetic components of PA and brain structures. However, such modelling requires considerably larger sample sizes than available in our study and the twin pairs should be unselected with respect to PA. Another approach to address the degree of genetic overlap is to use large population-based genome-wide association studies such as the UK Biobank, and on the other hand, apply Mendelian randomization analyses (Bowden and Holmes, 2019) to test for causality. Current estimates of the fraction of variance accounted for in self-reported PA and objectively measured PA in independent samples by polygenic risk scores derived from the UK Biobank are very small (Kujala et al., 2020). These latter approaches both require very large data sets.

3.2. Electrophysiological studies of brain function in twins associated with PA

Sensitive assessments of supple changes in normal cortical function are possible to register noninvasively with electroencephalography (EEG) and evoked potentials (EP). Mismatch negativity (MMN), an EP component detecting automatically changes in sensory environment, has been extensively studied in auditory and visual domains (Winkler and Czigler, 2012). MMN abnormality in the auditory domain has been closely associated with changes and decline in cognitive abilities in normal aging and neurological and neuropsychiatric illnesses (Näätänen et al., 2011). Although relatively less studied in somatosensory domain, the MMN has shown potential to serve as a reliable objective measure for sensory-cognitive functions across sensory domains (Strömmer et al., 2014). An important benefit of MMN paradigm is that it detects potential functional brain plasticity unrelated to conscious cognitive tasks (Strömmer et al., 2014). We have utilized this cortical involuntary change-detection system in somatosensory (Tarkka et al., 2016; Hautasaari et al., 2017) and visual (Pesonen et al., 2017) domains in PA discordant co-twin study design.

Somatosensory (SMMR) and visual mismatch responses (VMMR) were registered with continuous EEG in our sample of 22 MZ pairs (44 individuals) discordant in PA. SMMR was elicited by weak electrical stimuli delivered in second and fifth fingers producing location mismatch in averaged deviant stimuli (Tarkka et al., 2016; Hautasaari et al., 2017). VMMR was elicited by two types of visual stimuli presented on the screen (Pesonen et al., 2017). Participants were asked to ignore electrical or visual stimuli but to attend to an engaging audio program. The visual stimuli were black bars, each presented in turn for 100 ms in a grey background. Bars of the standard stimuli were tilted 18° to the right and bars for deviant stimuli were tilted 18° to the left, stimuli applied similar to Astikainen et al. (2008). From 1000 stimuli 10% were deviant stimuli and they were averaged to form VMMR.

Unequivocal mismatch components in typical latency windows both in somatosensory and visual domains were obtained in all co-twins. SMMR deviant waveforms were analyzed using multiple dipole modeling giving estimates of the locations and strengths of active sources. SMMR differed between co-twins so that stronger SMMR activation was observed in the right hemisphere in postcentral and frontal medial gyri and in left superior temporal gyrus in the co-twin with less PA, compared with the co-twin with higher amount of PA (Hautasaari et al., 2017). This modulation in involuntary SMMR activation may reflect differences between co-twins in selective attention processing and sensory gating of task-irrelevant stimuli. In other words, the ascending somatosensory stimuli may have alerted more the less active co-twins. In the visual domain, a reverse phenomenon was observed, where the co-twin with less PA showed longer VMMR latencies in occipital region when compared with co-twin with more PA (Pesonen et al., 2017). VMMR latency demonstrates the time taken by the nerve activation to travel and be processed in the involved cortical network. We found that those with more PA had faster reaction and faster preconscious automatic processing for random deviant visual stimulus. This faster processing occurred in occipital cortex, where large part of the primary and secondary visual processing in this time period takes place. While the genetic determinants of PA in general are evidently multifactorial and their biological markers largely unknown, the above described discordant co-twin study design provided one way to address the role of genetic factors in research focusing on PA.

Together these twin studies show that more PA may expedite preconsciously processing of visual stimuli and, in somatosensory domain, improve selective attentional processing by dampening the strength of unattended deviant somatosensory signals. These observed modifications seem independent of genetics and shared environment.

4. Conclusions

The moderate-strength result from observational studies of PA and reduction in cognitive decline has not been straightforwardly replicated in randomized controlled trials and genetically informed studies. The moderate-strength association may be partly explained by genetics and shared environmental factors.

While brain imaging studies on PA’s effects on macro-scale brain structure show inconsistent results, an increase in GM volumes in striatal, prefrontal, and hippocampal regions owing to PA has been
shown in MZ co-twin design. Additionally, more active MZ co-twins have shown differing automatic deviance-detection processes in brain regions involved with sensorimotor, visual and memory functions in electrophysiological studies. Currently there is very limited literature on genetically informed brain imaging addressing the relationship of PA and structural and/or functional brain imaging. It is not possible to draw comprehensive conclusions other than that the young healthy brain may show modulation associated with long-term PA. It would be important to understand how differences in PA influence the course of healthy or pathological cognitive aging. This knowledge would aid in discovering the most beneficial preventive measures. The purpose of studying twins, who are long-term discordant for PA, is to try to reduce the inevitable contribution of genetic selection in studies in attempts to explain the factors which mediate the benefits obtained from PA.

Acknowledgements

Authors wish to thank the participants of the studies. P. I-M. has been supported by Juho Vainio Foundation, Institute for Molecular Medicine Finland, University of Helsinki and HUS Medical Imaging Center. JK has been supported by the Academy of Finland (grants 308248 in 2017 and 312073 in 2018). PH has been supported by the Jenny and Antti Wihuri Foundation, Helsinki, Finland (grants 0170073 in 2017 and 00180080 in 2018). UK has been supported by FITTAFWIN: META-PREDICT (within the European Union Seventh Framework Program, HEALTH-F2-2012-277936 and the Finnish Ministry of Education and Culture (OKM 56/626/2013), Helsinki, Finland.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.neurobiol.2020.04.015.

References


