Examining the relationship between physical activity and cognitive function using a large population cohort

Thomas Campbell, M.A. Hons, MSc

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

July 2021

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Foreword

The primary aims of this project were completed largely in line with what was planned. However, because of disruption due to the Covid-19 pandemic, certain aspects were affected. Problems with remote access IT required for analysing a large dataset meant work was delayed by 4-6 weeks. This limited any capacity to complete additional analyses that might otherwise have been considered. There was also less opportunity to engage with the wider UK Biobank research group, limiting the scope for their collaborative input into aspects of the project that was initially planned.

Acknowledgements

I would like to thank my main supervisor Dr Breda Cullen for her erudite and responsive supervision throughout this thesis, and Dr Karl Ferguson for his specific guidance on all things DAG. Additional thanks to Dr Joey Ward and Dr Carlos Celis-Morales for their contributions to data coding.

I am deeply grateful to Heather and my parents for their support throughout the past three years, in the many forms this has taken.

My friends, clinical supervisors and colleagues have all contributed to my being able to complete this project, so thanks to each of you.

I appreciate the work of the entire team of lecturers, admin staff and tutors on the Glasgow DClinPsy programme, in providing me with support, skills and prospects for my career.

Cheers to inanimate providers of support such as bicycles, climbing walls and Mölkky pins, for giving me relief along the way.

Finally, I would like to acknowledge my brother, Hugh, who, though he is no longer with us, always plays an important role in my life.

Chapter 1: Systematic Review

Constructing a graphical model of the relationship between physical activity and cognitive function based on a systematic review of prospective evidence

Prepared in accordance with the author requirements for Journal of Neuropsychology [Appendix 1.1, p.91]

Systematic review abstract

Elucidating the factors that contribute to healthy ageing is an important research goal. Physical activity (PA) has been associated with benefits for cognitive function (CF). However, most of this evidence comes from longitudinal cohort studies which, in the absence of experimental design, have limited scope to make causal inferences regarding observed relationships. This review aimed to utilise recent methodological developments allowing researchers to formulate and answer stronger causal questions using observational data, by following a best-practice method for synthesizing evidence to produce a graphical causal model known as a directed acyclic graph (DAG). Following a search of 3 databases (EMBASE, MEDLINE and PsycINFO), 21 observational studies on the PA-CF relationship were reviewed and their methodological quality, characteristics, and key findings were summarised. The outcomes of interest were the covariates and modelling practices employed in each study. The reported covariates were synthesised against a set of criteria to determine their role in the DAG as confounders or mediators of the PA-CF relationship. Every included study had some areas of methodological weakness. The resulting DAG included a wide range of biopsychosocial covariates spanning the entire life-course and indicated potential intermediate pathways between PA and CF via structural brain health. Strengths, limitations and implications of this review for modelling decisions are discussed, prior to the model being taken forward to inform an empirical analysis using data from the UK Biobank cohort.

Word count: 228

SR Key words: Physical activity; cognitive function; directed acyclic graph; healthy ageing

Introduction

Background

The term cognitive function (CF) describes the set of mental abilities that enable the acquisition and use of knowledge and skills throughout life. Humans vary in these abilities and lower CF is associated with a range of negative health outcomes (Batty, Deary, & Gottfredson, 2007). Reduction in certain aspects of CF occurs as part of normal aging (Harada, Love, & Triebel, 2013), but decline beyond the normal range is a feature of clinical presentations, from mild cognitive impairment (MCI) through to Alzheimer's disease and other dementias (Gauthier et al., 2006; Husain & Schott, 2016). Due to the aging population, the prevalence of conditions involving cognitive impairment is expected to rise and place increasing burden on society (Nichols et al., 2019). Elucidating the factors associated with CF is thus an important task for researchers.

Physical activity (PA) is defined as *"... any bodily movement produced by skeletal muscles that requires energy expenditure"* (Casperson et. al., 1985, p.126) and thus, as a concept, includes everyday activities such as DIY, walking and shopping as well as purposeful exercise and sport. PA has emerged as a modifiable risk factor associated with CF and neurodegenerative conditions (Blondell, Hammersley-Mather, & Veerman, 2014; Sofi et al., 2011). There are several pathways by which PA may affect CF. At the cellular level PA appears to facilitate neurogenesis, synaptogenesis and angiogenesis (Lista & Sorrentino, 2010) which may produce structural changes in grey matter volume (Erickson et al., 2019) and white matter integrity (Sexton et al., 2016). Other intermediate pathways between PA and CF may include behavioural and psychological mediators such as sleep and mood (Stillman, Cohen, Lehman, & Erickson, 2016).

The gold standard of evidence in health science research is an experimental design such as the randomised controlled trial (RCT), and there is evidence from such trials demonstrating effects of targeted exercise on CF (Northey, Cherbuin, Pumpa, Smee, & Rattray, 2018). However, the broader conceptualisation of PA is less feasible to study experimentally and so most of the evidence comes from longitudinal cohort studies

which track people's levels of PA 'in the wild' and estimate associations with various outcomes. Due to the observational design of such studies, researchers utilising these paradigms have traditionally been cautious about making causal claims from their findings (Rohrer, 2018). This is a position which is prudent, yet at odds with the aim of research to explain phenomena. Furthermore, such designs commonly utilise statistical adjustment strategies to improve their estimates without careful consideration of the assumptions behind, and implications of, these decisions, and thus may unwittingly introduce more bias into their models (Hernán, 2018). Examples of adjustment errors include failing to adjust for a confounder, over-adjustment for a mediator and inappropriate adjustment for a collider. In response to these limitations, researchers seeking to formulate and answer stronger causal questions using observational data have laid out an approach utilising a graphical tool known as a directed acyclic graph (DAG) (Pearl, Glymour, & Jewell, 2016). Constructing a DAG requires researchers to explicitly state the effect they are interested in (the 'estimand'), and then lay out their assumptions regarding relationships between the exposure, the outcome, and the relevant covariates. Through application of an algorithm to the specified model, a DAG can identify the necessary adjustment decisions required for causal interpretation of an estimand (see figure 1 for a more detailed description of DAGs in theory and practice).

Of course, any DAG is only as good as the knowledge put into it by the researcher. However there has been a lack of guidance on how to produce DAGs in a robust way, until the recent publication of a methodological protocol on how to synthesise evidence to construct DAGs (Ferguson et al., 2020) and recommendations on how to present them (Tennant et al., 2019). Applying these methodological developments in the context of the PA-CF literature, can help identify key confounders and mediators and guide modelling decisions.

Objectives

This review aimed to utilise the recently developed method for constructing DAGs: Evidence Synthesis for Constructing Directed Acyclic Graphs (ESC-DAGs; Ferguson et al., 2020), applied to observational studies of the association between PA and CF. The

resulting DAG was then used to inform an empirical analysis of UK Biobank data, presented in Chapter two of this thesis. The purpose of this review was not to investigate the magnitude of associations between PA and CF; rather, it was to systematically investigate the factors that covary with PA and CF, in order to inform the construction of a causal DAG depicting the confounders and mediators of the relationship between PA and CF. Specifically, this review aimed to answer the following questions:

(a) What clinical and sociodemographic factors, and health behaviours, do the included studies report as being associated with cognitive function/impairment/decline and/or physical activity?

(b) How do the studies vary with regards to: the covariates included, how these are adjusted for, and the conclusions of the study?

Figure (box) 1: DAGS – the basics

Directed acyclic graphs (DAGs) represent a useful tool for researchers to lay out their theories and assumptions regarding the causal relationship between variables. A definition of causality that is compatible with DAGs has been posited as "A variable X is a cause of Y if Y in any way relies on X for its value" (Pearl et al., 2016). In practice this means DAGs are non-parametric in that they make no assumptions regarding the functional form of the relationships (e.g. linearity), the strength of relationships, or the causative direction (e.g. harmful or protective). The statistical software used in this thesis, DAGitty (Textor, Hardt, & Knüppel, 2011), applies an algorithm, known as the d-separation criterion (Pearl, 2009), which allows one to establish which variables need to be adjusted for, and which to leave unadjusted, given the researcher's assumptions regarding the relationships between each pair of variables. The required variables are referred to as the minimally sufficient adjustment set.

A DAG depicts conceptual variables as nodes, and putative causal effects between variables as arrows between nodes. The absence of an arrow between nodes encodes the strong assumption of no causal relationship for any member of the population. For a DAG to be complete and have a causal interpretation, all shared causes of all pairs of nodes must be depicted (these can be labelled as unknown factors if necessary). A sequence of nodes connected by arrows is a path, and paths must be acyclic (no path can feed back into itself). This acyclicity encodes the assumption that any variable, at a given point in time, cannot cause itself (although the same variable at different time-points, depicted as separate nodes, may cause one another). In terms of applying these principles to examples relevant to the present review, consider the simplest relationship depicted within a DAG, applied to the exposure and outcome of interest (figure 1.1). Note that DAGitty depicts the exposure variable as green and the outcome as blue with the symbol ▎, and the presence of an unbiased path with a green arrow. Figure 1.1 encodes the following assumptions:

1) The value of CF (i.e., whether it is low or high) depends in some way on the value of PA, for at least some members of the population.
2) There are no other variables which have direct associations with both PA and CF (strong assumption of no shared causes).

Figure 1.1: DAG representing the posited causal relationship between PA and CF

Naturally, such simplistic DAGs are unlikely to exist in relation to complex phenomena with multiple biopsychosocial causes. As more variables are integrated into a DAG, the assumptions made regarding the relationships between them determine the role each variable plays in transmitting the effect of interest. Three types of variable that should be considered in models are confounders (shared causes of the exposure and outcome which must be depicted in order for a DAG to have a causal interpretation), mediators (variables which transmit some of the effect of interest via an indirect pathway and are not mandatory to depict unless the researcher is interested in conducting mediation analyses), and colliders (a variable that is a shared outcome of exposure and outcome). Researchers may also be interested in moderators (also known as effect measure modifiers), which are variables that interact with each other to affect the outcome. Since DAGs are non-parametric objects, moderators are simply depicted in the same way as confounders, but the researcher can model these statistically using interaction terms when they translate the DAG into a statistical model specification for their estimand of interest (e.g. regression equation).

Simple examples of how confounders (figure 1.2), mediators (figure 1.3) and colliders (figure 1.4) would be depicted within a DAG are shown below.

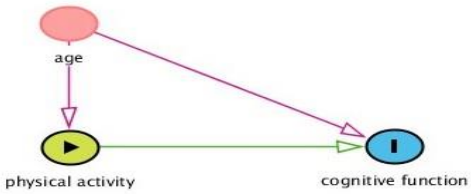


Figure 1.2: DAG depicting age as a confounder of the PA-CF relationship.

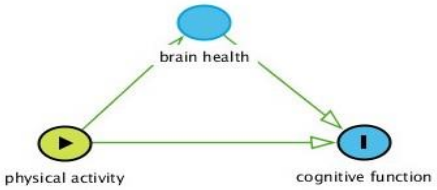


Figure 1.3: DAG depicting brain health as a mediator of the PA-CF relationship.

In figure 1.2 the values of both PA and CF are assumed to rely in some way on the value of age. Age opens a biasing pathway between PA and CF, sometimes referred to as a backdoor path in the DAG literature. Age is therefore depicted as a confounder which must be adjusted for in order to 'unbias' any estimate of PA's effect on CF. Note that the DAGitty software depicts confounders as pink nodes, and biased pathways as pink arrows. In figure 1.3 the value of brain health is assumed to rely in some way on PA, and the value of CF to rely in some way on brain health. Thus, as brain health is causally 'down-stream' of the exposure, it is represented as a mediator, and part of the total effect of interest. Mediators should thus not be adjusted for in estimating the effect total of PA on CF. Note that DAGitty represents mediators as blue nodes and unbiased paths between mediators and outcome as green arrows.

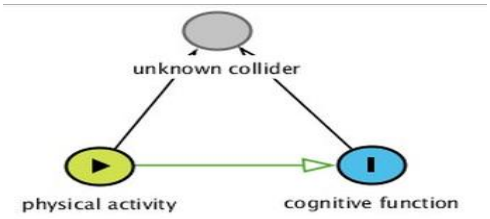


Figure 1.4: DAG depicting an unknown variable mutually caused by PA and CF (a collider)

In figure 1.4 a variable whose value is assumed to rely in some way on both PA and CF, is depicted as a collider. Importantly, a collider's structural position blocks spurious association from flowing between exposure and outcome via this path and should thus not be adjusted for. When colliders are, often unintentionally, adjusted for, the backdoor path is opened, and can bias the exposure-outcome relationship, known as collider conditioning (Elwert & Winship, 2014). This can often be a problem in research when participation in the study, or missing data within the study, is related to exposure and outcome.

Figure 1.5 illustrates a simple DAG with each type of variable represented. In this case DAGitty indicates that the minimally sufficient adjustment set required to estimate the total effect of PA on CF contains age and genetics (confounders), whilst brain health is part of the effect of interest (a mediator) and should be left unadjusted.

Possible colliders are represented by a single grey node, and DAGitty indicates that these should not be adjusted for.

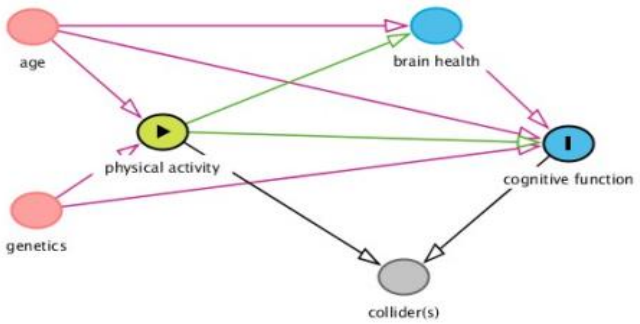


Figure 1.5: Simplified DAG depicting hypothetical confounders, mediators, and colliders.

Method

Protocol and registration

A protocol for this review was written in accordance with the COSMOS-E guidance on conducting systematic reviews of observational studies of etiology (Dekkers et al., 2019). This was registered on the Open Science Framework registry on 04/09/2020 and is available from https://osf.io/wuycz/. This review has been written in accordance with COSMOS-E guidance so far as is possible, given the novel method of synthesis applied, and incorporates features of PRISMA guidance (Page et al., 2021) where appropriate.

Eligibility Criteria

- *Condition being studied:* Cognitive function as demonstrated by performance on objective cognitive tests. Where the outcome was measured categorically as impaired versus unimpaired, the threshold for impairment was as defined by study authors but would at a minimum require performance to be one standard deviation below the mean of a healthy comparison group.
- *Types of studies:* Longitudinal cohort studies of prospective design were included. This design reflects the Biobank cohort resource which will be used for subsequent empirical analysis (Chapter Two).
- *Population:* Community dwelling adults, free of cognitive impairment before the period of exposure to physical activity began and aged ≥45 to <80 at baseline. The upper limit of <80 serves to reduce the likelihood of undetected

pre-clinical cognitive decline at baseline, which would be expected to be higher in older samples. The lower limit serves to ensure that included samples are at least 'middle aged' at baseline, which reflects the Biobank cohort (Sudlow et al., 2015). Therefore, studies whose entire sample was <45 or ≥80, at baseline, were excluded. Where the sample's range included ages both within and outside the ≥45 and <80 range, the study was only included if the mean age of the sample at baseline was between ≥45 and <80.

- *Exposure:* Level of physical activity, measured using either self-report questionnaires such as the International Physical Activity Questionnaire (Craig et al., 2003) or objective measures such as actigraphy. Where the exposure was measured categorically as active versus inactive the threshold for inactivity was as defined by study authors (e.g. not meeting World Health Organisation guidance for weekly PA).
- *Comparator:* Given that the exposure (risk factor) is low physical activity, the unexposed group is those who engaged in greater levels, as defined by the study authors.
- *Outcomes:* Given the nature of the present review questions, the outcomes were the covariates included in each study (clinical and sociodemographic factors, and health behaviours), and how these were treated in the analysis (e.g. as confounders, mediators or moderators).

Search procedures

The following databases were searched on 28/07/2020: Medline on the Ovid platform; Embase on the Ovid platform and PsycINFO on the EBSCOhost (EBSCO) platform. The search was restricted to English language publications between 01/01/2005 and the date of search. This date range was deemed appropriate as visual inspection of a histogram plotting the frequency of relevant articles by year indicated that a large majority of results were from after 2005. The sensitivity of the search strategy was tested by checking whether key papers known to be relevant were captured, and search terms were modified accordingly. Additional articles were identified by hand-searching reference lists of relevant papers. See appendix 1.2 (p.97) for the full search

strategy including search terms and appendix 1.3 (p.98) for the test of search sensitivity.

Search results were exported to EndNote X9 and titles and abstracts were screened by the lead author, using an eligibility checklist developed for this review (appendix 1.4). Articles which did not meet eligibility criteria were discarded. A second researcher (JW) screened a subset (n=100) of titles and abstracts. Inter-rater agreement was 84% (*k =.56)* indicating substantial consistency. Disagreements were resolved by discussion. Full texts for remaining articles were retrieved and screened by the lead author. A subset (n = 20) was screened by JW with agreement at 70% (*k=.29),* indicating fair consistency. As several disagreements occurred for the same reason these were discussed with a senior author (BC) and an amendment to the checklist was made providing clearer instructions (appendix 1.4, p.99). A further subset of 5 was screened by JW with 100% agreement.

Data extraction

The final set of eligible papers was ranked according to the number of covariates included. This method was chosen for efficiency, as it allowed the complete array of covariates amongst eligible articles to be captured by the minimum set of studies. Therefore, of the 60 eligible articles, 21 were taken forward for synthesis. Ranking articles according to their quality rating was also considered but deemed impractical. Data extraction from the articles included in the synthesis included the following variables: definition and measurement of PA and CF, covariates included in adjusted models and modelling practices employed (see appendix 1.5, p.102 for the data extraction form). Data extraction was carried out by the lead researcher, with a second researcher (JW) checking a subset of these (n=10) for accuracy and completeness.

Risk of bias/quality rating

COSMOS-E guidance (Dekkers et al., 2019) recommends that confounding bias, selection bias and information bias are assessed when evaluating observational studies of etiology. To ensure these were appropriately assessed, each included study was evaluated by the first author for methodological quality and bias using a critical

appraisal tool for prospective studies – The Joanna Briggs Institute (JBI) Checklist for Cohort Studies (Moola et al., 2017). In addition to this tool, item A2 from the Critical Appraisal Skills Programme (CASP) Cohort Study Checklist (2018) was included, to ensure that selection bias was directly assessed. In line with COSMOS-E guidance, no overall risk of bias score was assigned to each study; rather, the ratings for each item are presented transparently so that areas of strength or concern can be observed within and between studies (appendix 1.6, p.103). A second researcher (JW) independently rated a random sample (n=5) of the papers for comparison (90% inter-rater agreement). Rating discrepancies were resolved by consensus.

Synthesis strategy

In the context of the present review questions, the aim of data synthesis was to elucidate and represent the causal structure of the relationships between PA, CF and covariates. Tables and narrative summaries were used to describe the range of covariates and modelling practices within the included articles. These were then synthesised to produce a causal diagram using the ESC-DAG method (Ferguson et al., 2020), involving translating the conclusions of each included study into an individual DAG (see figure 2 for an illustration). All implied relationships were extracted into a combined index, and then all possible pairwise relationships between variables were assessed against a set of causal criteria. The resulting fully integrated DAG was then subjected to expert opinion (KF and BC), with any further variables not captured by the included literature integrated to the model by consensus.

Figure (box) 2: A demonstration of ESC-DAGs

The ESC-DAGs process is described in detail by Ferguson et al. (2020). The method is applied below to a hypothetical study examining the effects of PA on CF with a small number of covariates (APOE-e4 genotype, age, brain health measured by MRI and smoking status).

Stage 1: Mapping

Purpose: To apply graph theory to the conclusions of each study. This creates an 'implied graph' (IG) which acts as a transparent structural template for translation into a DAG.

1. Outcome variable of interest is set as DAG outcome(s).
2. Exposure variable(s) of interest is set as DAG exposure(s).
3. An arrow is drawn originating from the exposure(s), terminating at the outcome(s).
4. All control variables are entered as unassigned variables.

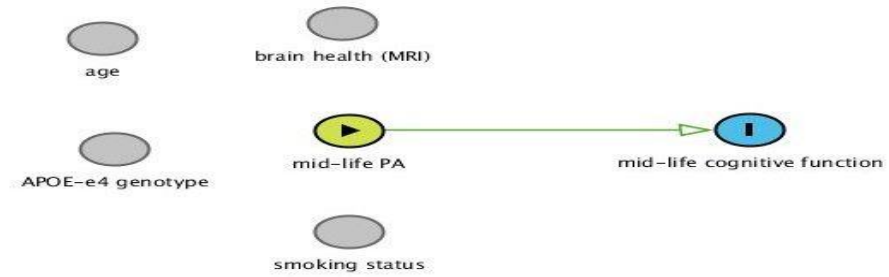


Figure 2.1: ESC-DAGs mapping stage, steps 1-4

(Box 2 continued)

5. An arrow is drawn originating from each control variable to the exposure(s) and outcome(s).

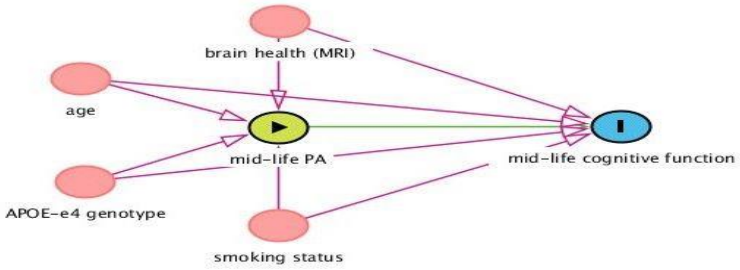


Figure 2.2: ESC-DAGs mapping step 5

6. Mediators, instrumental variables etc. are mapped as per the study's conclusions and noted.
7. The IG is saturated by drawing arrows between all confounders (direction does not matter until the translation stage) and all arrows are extracted to the main index. The recombination process (below) can be performed at this stage to help simplify an overly complex IG).

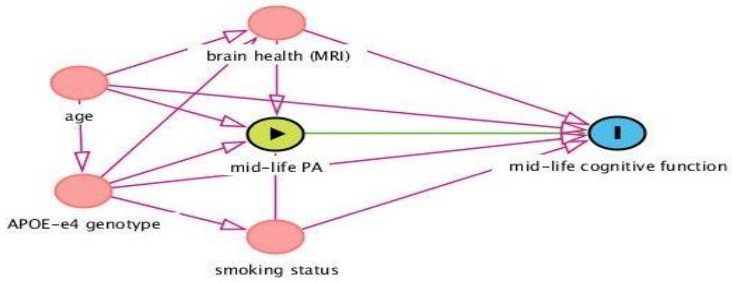


Figure 2.3: ESC-DAGs mapping steps 6 & 7

Stage 2: Translation

Purpose: To apply causal theory to each relationship in the IG. This creates the DAG for the study. Each relationship (arrow) in the IG is assessed under sequential causal criteria and a counterfactual thought experiment.

The posited arrow and its reverse are both assessed. Arrows may be retained as posited, reversed, or as bi-directional. Otherwise, they are deleted. All retained arrows are entered into an index of relationships (figure 2.4).

(Box 2 continued)

1. Temporality—can the posited cause precede effect? (If 'yes', proceed to next criterion. If not, assess reverse relationship.)
2. Face-validity—is the posited relationship plausible? (If 'yes', proceed to next criterion. If not, assess reverse relationship.)
3. Recourse to theory—is the posited relationship supported by theory? (Always proceed to the counterfactual thought experiment.)

Study	Variables			Casual Criteria					
				Posited direction			Reverse direction		
Hypothetical, 2021	Variable	Measurement	Posited cause-effect	*Temporally valid*	*Face valid*	*Theoretical support*	*Temporally valid*	*Face valid*	*Theoretical support*
	APOE-e4	Genetic sequencing	APOE-e4 -> CF	yes	yes	yes, Vergehse et.al., 2011	no	n/a	n/a
			APOE-e4 -> PA	yes	yes	no	no	n/a	n/a
	Age	Self-reported age	Age -> CF	yes	yes	Yes, Salthouse, 2009	no	n/a	n/a
			Age- -> PA	yes	yes	no	no	n/a	n/a
	Brain health	MRI - grey matter volume	Brain health -> CF	yes	yes	Yes, Deary et.al, 2010	yes	no	n/a
			Brain health -> PA	yes	yes	Yes, Sallis et.al., 2000	yes	yes	yes, Erickson et.al. 2011
	Midlife CF	Cognitive battery	N/A						
	Midlife PA	IPAQ self-report questionnaire	N/A						
	Smoking status	Self-reported	Smoking -> CF	yes	yes	Yes, Sabia et.al, 2008	yes	yes	no
			Smoking -> PA	yes	yes	yes	yes	yes	Yes, Heydari, et.al, 2015

Figure 2.4: ESC-DAGs translation steps 1-3

4. Counterfactual thought experiment—is the posited relationship supported by a systematic thought experiment to explicitly draw out the implications of the posited assumption (once completed, always assess the reverse relationship unless already assessed).

Study	Variables			Counterfactual thought experiment					
				Posited Direction			Reverse direction		
Hypothetical, 2021	Variable	Measurement	Posited cause-effect	*Posited cause condition 1*	*Posited cause condition 2*	*Hypothesised outcome*	*Posited cause condition 1*	*Posited cause condition 2*	*Hypothesised outcome*
	APOE-e4	Genetic sequencing	APOE-e4 -> CF	All sample have high risk allele	None of sample have high risk allele	Reduced CF in condition 1	n/a	n/a	n/a

Figure 2.5: ESC-DAGs translation step 4.

5. Conclusions: deciding whether to retain, reverse or remove the arrow between two nodes.

(Box 2 continued)

	A	B	C	D	E	F
1	Study	Variables			Conclusions	
2	Hypothetical, 2021					
3		**Variable**	**Measurement**	**Posited cause-effect**	***Outcome***	***Direction***
4		**APOE-e4**	Genetic sequencing	APOE-e4 -> CF	retain posited edge	unidirectional
5				APOE-e4 -> PA	retain posited edge	unidirectional
6		**Age**	Self-reported age	Age -> CF	retain posited edge	unidirectional
7				Age- -> PA	retain posited edge	unidirectional
8		**Brain health**	MRI - grey matter volume	Brain health -> CF	retain posited edge	unidirectional
9				Brain health -> PA	reverse posited edge	unidirectional
10		**Midlife CF**	Cognitive battery	N/A	N/A	N/A
11		**Midlife PA**	IPAQ self-report questionnaire	N/A	N/A	N/A
12		**Smoking status**	Self-reported	Smoking -> CF	retain posited edge	unidirectional
13				Smoking -> PA	retain posited edge	bidirectional
14						
15						

Figure 2.6: ESC-DAGs translation step 5.
This process is repeated for all relevant studies that the researcher aims to integrate into the final conceptual model.

Stage 3: Integration 1 (synthesis)

Purpose: To combine the translated DAGs into one by synthesising all indexed relationships.
1. A new DAG is created to serve as the integrated DAG (I-DAG).
2. The focal relationship is added to the I-DAG (as per mapping steps 1–3).
3. Each indexed arrow pertaining to the focal relationship (including its corresponding node) is added to the diagram.
4. Each indexed arrow pertaining to other nodes is added (e.g. between confounders).
5. Conceptually similar nodes should be grouped together in virtual space to aid the recombination process.

Stage 4: Integration 2 (recombination)

Purpose: To combine nodes for either practical reasons (i.e., to reduce complexity) or substantive reasons (i.e., to establish consistency).
1. Is there theoretical support for combining two variables/nodes?
2. Do the conceptually related nodes have similar inputs and outputs (i.e., do they 'send to' and 'receive from' the same nodes)?

Results

Study selection

The study identification process is illustrated in Figure 3. The database search yielded 5,898 results, and one additional paper was identified by hand-searching. The test of search sensitivity indicated that all 10 relevant papers were captured. A total of 60

studies were eligible. Given that the aim of this review was to identify and synthesise the range of covariates identified in primary studies, papers were synthesised in descending order by total number of covariates until the point that including further studies would only duplicate existing variables. To implement this process the first author counted and ranked the number of covariates from all 60 studies. This resulted in a total count of 73 covariates that were captured by 21 papers, which were taken forward for synthesis. After recombining different measures of similar constructs, the total number of conceptual variables was reduced to 56 (e.g., rather than representing BMI and waist circumference as separate nodes these were recombined into a single conceptual variable 'adiposity'). Table 1 describes the key characteristics of the studies including age and size of samples, how PA and CF were defined and measured, and the key findings. Table 2 reports the covariates and modelling practices within each study.

Risk of bias/quality results

Full risk of bias/quality ratings are displayed in appendix 1.6. All of the included studies suffered from methodological limitations in at least one area. Most commonly this pertained to the recruitment of non-representative samples, introducing the possibility of selection bias, incomplete follow-up which compromises the internal validity of findings, and quality of the measurement of exposure, with many studies using unvalidated self-report measures to capture PA. For all but two studies there were concerns regarding differences between the exposed and unexposed groups on relevant covariate measures, which indicates high risk confounding bias. However, this was addressed by all studies through identification of confounders, and by implementing strategies to adjust for them (all studies). Another area of strength was in the outcome measures used, with all but one reporting assessments of CF using valid and reliable measures.

Exposure measures and classification

Most studies used self-report questionnaires to measure PA (only three used accelerometery). Studies most commonly divided the sample (e.g., into quartiles or tertiles) according the distribution of PA within the sample, though six studies entered

PA as a continuous variable. Only two studies used objective criteria (WHO guidance) to dichotomise their sample. Six studies entered PA as a continuous measure.).

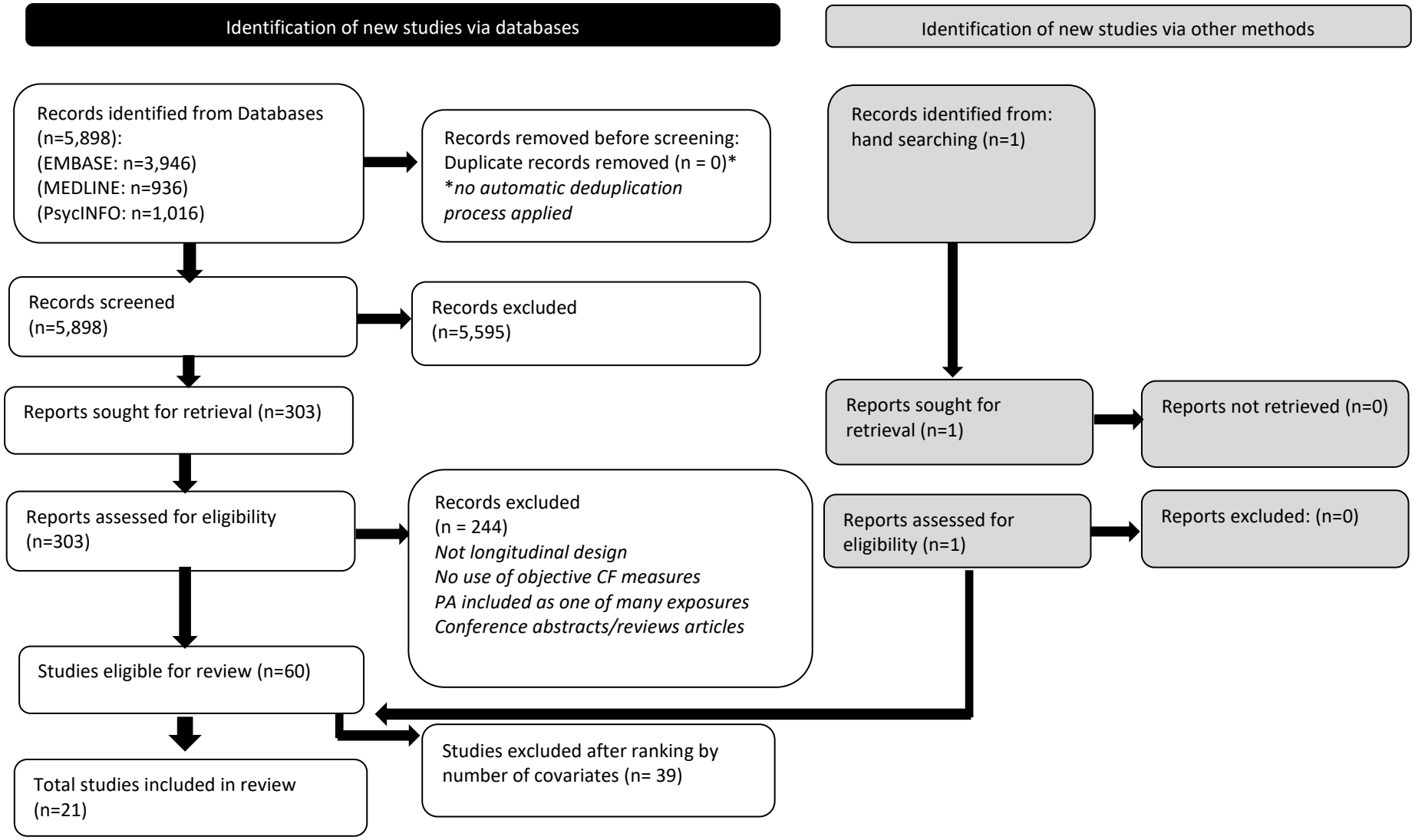


Figure 3: PRISMA flow diagram

Outcome measures and classification

The most common measures of CF were cognitive screening tools such as the Mini Mental State Examination (MMSE) which capture global CF. In some cases, batteries of standardised tests such as the Symbol Digit Modalities Test (SDMT) and the Stroop test, or bespoke (unstandardised) mental tasks, were used to measure separate cognitive domains. Other outcome measures included dementia diagnosis where reference was made to cognitive testing and impairment thresholds, but specific measures were not reported. Studies varied between modelling CF as a continuous variable, or as categorical, (i.e., impaired vs unimpaired). Where CF was treated as continuous, methods included using the raw scores from individual tests, or total scores from a screening tool. This was achieved in some cases by converting raw scores to z-scores centred around the sample mean and taking the mean of the z-scores to indicate overall performance. Studies that used categorical definitions of impairment used existing criteria such as scores below a threshold on a screening measure (e.g., scores of <80 on 3MS), deviations from the sample mean of certain magnitude (e.g., z-scores at least 1.5 away from the mean), or reductions in CF of certain magnitude between baseline and follow-up (e.g., decline of >3 points on the Blessed Test).

Key findings

Of the 21 included studies, 18 reported some form of significant protective association between PA and CF, and three reported no significant association. Of these three, two were distinctive in terms of having younger baseline samples and longer follow-up periods than other studies (Morgan et al., 2012; Sabia et al., 2017). Furthermore, one study (which reported an overall protective association between PA and CF) found no association in a sub-analysis of participants who were retained longer than 10 years (Tan et al., 2017). One explanation explored by the authors to account for these anomalous findings, was that studies with older baseline samples may be at risk of reverse causation, whereby participants may already have preclinical cognitive impairments, and their lower PA is a symptom, rather than a cause, of these cognitive changes.

Adjustment strategies

The most common modelling technique was multiple regression, with confounders identified by statistically significant associations with the exposure and/or outcome. No study employed causal inference methods such as use of a DAG to guide modelling decisions. It was typical for authors to speculate about mediating mechanisms, based on significant regression coefficients, without formally testing these hypotheses through mediation analysis. In some instances, PA was treated as a moderator of another risk factor for CF (e.g., genetic risk, or sodium intake). Reduction of cardiovascular risk was often discussed as an intermediate pathway by which PA affects CF.

Covariates reported by included studies

Genetic

The most common genetic covariate was APOEε4, which was included in adjusted models by six studies. Being a carrier of the ε4 allele, as opposed to the ε3 and ε2 alleles, is associated with higher risk of Alzheimer's disease and cognitive decline (Liu, Kanekiyo, Xu, & Bu, 2013). The only other genetic variables were both included in the same study. These were variants of Insulin Degrading Enzyme and Brain Derived Neurotrophic Factor genes, which have both been associated with brain health and CF.

Sociodemographic

The most common sociodemographic variables included were age (all studies) and education (sixteen studies). Other commonly included variables were socio-economic status, ethnicity and marital status. One study included acculturation and one included measures of social support.

Health behaviours

The most common health behaviours included related to alcohol intake (ten studies) and smoking status (nine studies). Other variables were related to diet, including sodium intake and use of vitamins.

Table 1: Characteristics of the included studies

Study	Country	Age at baseline	N at baseline	Follow-up time[a]	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings
(Barha et al., 2020)	USA	M= 73.6	2,873	10 years	Self-reported total minutes walking per week	Participation in walking at least 10 X in last year	EF and PS: DSST Global Cognition: 3MS	Scores <80 on 3MS	Positive association between PA and global CF in both sexes. Positive associations between PA and EF and PS in females only. Positive association between PA and left dlPFC in females only. Negative association between PA and left hippocampal volume in females; positive in males.
(Etgen et al., 2010)	Germany	M = 71.2	3,903	Median = 778 days	Self-report questionnaire on frequency of strenuous activities	No activity Moderate activity: < 3 times/week High activity: ≥ 3 times/week	Global cognition: 6-CIT	Scores > 7	Negative association between PA at baseline and incident cognitive impairment at follow-up.
(Fiocco et al., 2012)	Canada	M = 74.2	1,793	3 years	Self-report: Physical Activity for Elderly Scale (PASE)	High or low activity based on median split PASE score	Global cognition: 3MS	Scores <80 on 3MS	Negative association between sodium intake and CF in low PA group only
(Gow, Pattie, & Deary, 2017)	UK	M = 79	550	11 years	Retrospective PA: bespoke self-report measure rating general PA on 6 point to scale at three age periods (20-35; 40-55; 60-75) Contemporaneous PA (recorded at 79, 83, 87 and 90): bespoke self-report measure rating frequency of various leisure activities including PA	Retrospective PA: Exploratory factor analysis conducted, with standardised residuals used to define an activity score for leisure activity and PA for each age category Contemporaneous PA: as above	Scores on standardised tests (phonemic verbal fluency, Logical Memory and Raven's progressive matrices) were used to extract a general latent cognitive ability (g) factor, which was used as the dependent variable	Cognitive change, entered as a continuous variable	Negative association between PA and cognitive decline.

Study	Country	Age at baseline	N at baseline	Follow-up time[a]	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings
(Halloway, Wilbur, Schoeny, & Barnes, 2017)	USA	M = 63 for those who attended follow-up M= 67.1 for those who dropped out	174	M= 5.2 years	Self-report: lifestyle PA frequency and intensity, measured using the Community Healthy Activities Model Program for Seniors Questionnaire. Responses were assigned MET values to calculate total amount of moderate and vigorous PA Objective: Accelerometer data used to calculate mean daily minutes of light, moderate-vigorous and total PA.	For self-report and objective measures PA was quantified in terms of amount of light, moderate - vigorous and total PA, per week (self-report) or per day (objective)	Battery of cognitive tests measuring PS, semantic memory and episodic memory. Scores were expressed as Z scores and these were used to create composite scores for global CF and each subdomain.	Cognitive decline was defined as a drop of CF of >0.5 SDs from the mean rate of decline	Negative association between self-reported light PA and risk of episodic memory decline. Negative association between accelerometer measured moderate-vigorous PA and risk of semantic memory decline.
(Hamer, Terrera, & Demakakos, 2018)	UK	M= 65	10,652	10 years	Self-reported participation in mild, moderate and vigorous PA activities using a four point scale.	Categorised into four groups: inactive (no PA on weekly basis), only mild at least once per week, at least moderate but no vigorous once per week and any vigorous activity at least once per week.	Memory: a ten-word recall test. EF: a category fluency test.	Number of correctly recalled words and correctly named animals were used as continuous measures of CF.	Positive association between PA at baseline and preservation of memory and EF over 10 years in females, and in EF only for males
(Iso-Markku et al., 2016)	Finland	M = 49.1	3,050		Self-report questionnaire regarding volume and intensity of leisure-time PA and commute-based PA, used to calculate total PA expressed in MET, then divided into quintiles for the sample.	Separate measurements pertaining to different years were combined to reflect groups change in PA over time (being in the	Global cognition: TICS	Total score used as a continuous measure of global CF	Negative association between level of PA and risk of cognitive impairment.

Study	Country	Age at baseline	N at baseline	Follow-up time[a]	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings
						lowest quintile both years = 'persistently inactive'; moving out of lowest quintile = 'changed activity'; highest quintile both years = 'persistently active')			
(Larson et al., 2006)	USA	M = 73.2	1,740	M = 6.2 years	Self-report questionnaire regarding frequency of participation in a range of exercise for at least 15 minutes. Frequency of exercise was calculated as total number of such episodes per week.	Being active was defined as exercising at least three times per week.	Global cognition: CASI at baseline, and CASI plus dementia diagnostic assessment at follow-up. The outcome variable was incident dementia.	A CASI score of <86 was necessary for dementia diagnosis.	Negative association between exercising 3 or more times per week and incident dementias.
(Morgan et al., 2012)	UK	56	2,959	16 years	Self-report questionnaires regarding duration and frequency of work-related and leisure-time PA.	Both work-related PA and leisure time PA were divided into tertiles to define low, moderate and high activity groups.	Global cognition: CAMCOG, plus dementia diagnostic assessment for positive screens.	A CAMCOG score of <83 was necessary for dementia assessment.	No significant association between PA and dementia.
(Ogino, Manly, Schupf, Mayeux, & Gu, 2019)	USA	75.1	1,345		Current leisure time PA: self-report questionnaire regarding frequency and duration of activities of activities of various intensity (according to MET weights). Total time was used to calculate MET hours/week.	Current leisure time PA was categorised as low, moderate or high according to MET hours. Past leisure time PA was also categorised as	Clinical assessment for dementia/MCI which included neuropsychological battery (tests not reported).	At least 1.5 SDs below the mean on the battery plus positive dementia diagnosis	Negative association between both current and previous LTPA and risk of AD.

Study	Country	Age at baseline	N at baseline	Follow-up time[a]	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings
					Past leisure time PA was measured retrospectively using similar questions to estimate MET minutes for life periods 12-25, 26-50 and > 50 years old.	low, moderate and high based on tertiles.			
(Papenberg et al., 2016)	Sweden	71.8	555	6 years	Self-report questionnaire regarding duration and frequency of participation in activities categorised by intensity.	Three categories of PA created based on WHO guidance: inadequate (less than 3 episodes of light and/or moderate-vigorous PA per month); health-enhancing (light exercise at least several times per week); fitness-enhancing (moderate/intense exercise several times per week).	MMSE	MMSE change score over 6 years	PA had a protective interactive effect on negative association between inflammation markers, brain health and CF.
(Podewils et al., 2005)	USA	74.8	3,375	M = 5.2 years	Self-report questionnaire regarding frequency and duration of activities assigned MET weightings and used to estimate PA-related energy expenditure per week. An index of number of activities participated in ranging from 0-14 was also used as a continuous measure of PA frequency.	Energy expenditure and PA frequency used as continuous measures	Dementia diagnosis including administration of Modified MMSE (3MS), or TICS (if person did not receive clinical evaluation).	3MS scores below 80 or decline of 5 or more points. TICS score of below 28	Negative association between PA frequency and risk of all dementias

Study	Country	Age at baseline	N at baseline	Follow-up time[a]	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings
(Rabin et al., 2019)	USA	73.4	182	Median = 6 years	Objective: pedometer used to calculate mean steps per day	Number of steps per day	Global cognition (PACC)	Raw scores were z-transformed and averaged together for continuous measure of CF	Higher levels of PA attenuated the associations between AB burden and cognitive decline and neurodegeneration
(Rajan et al., 2015)	USA	73.2	7,742	M = 9.5 years	Self-report: questionnaire regarding frequency and duration of activities	Composite index created by summing the products of minutes in each activity, then split into quartiles: 0, 0.01-1.24, 1.25-3.99 and >4 hrs/week)	Global cognition (MMSE) Episodic memory EBT PS (SDMT)	Raw scores were z-transformed and averaged together for continuous measure of CF	Negative association between PA and rate of cognitive decline amongst racially white sample, but no association in racially black sample.
(Reas, Laughlin, Bergstrom, Kritz-Silverstein, & McEvoy, 2019)	USA	73.5	2,212	M = 5.9 years (active group); 8.4 (inactive group) Total sample M not reported	Self-report: questionnaire regarding frequency and duration of PA, in teens, 30s and concurrently at each assessment wave	Classified as active if engaging in any level of PA ≥ 3 times/week	Global cognition (MMSE) EF (Trails B; verbal & semantic fluency) Episodic memory (BSR) Measured at multiple time-points to allow for trajectories to be calculated	Each test entered as continuous measure of CF	Positive association between concurrent PA and better CF, which was age dependent (stronger effect at older ages). PA in early adulthood also augments this association.
(Rovio et al., 2005)	Finland	50.8 (active) 49.5 (inactive)	1,449	M = 20.7 (active) M = 21.3 (inactive)	Self-report: questionnaire regarding frequency and duration of PA	Dichotomous classification of active if engaging in any level of PA ≥ 2 times/week and	Global cognition (MMSE) and full clinical dementia assessment for positive screens.	Score of <25 on MMSE	Negative association between mid-life PA and risk of dementia

Study	Country	Age at baseline	N at baseline	Follow-up time[a]	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings
		Total sample M not reported		Total sample M not reported		inactive if <2 times/week	Incident dementia was the outcome variable		
(Sabia et al., 2017)	UK	44.8 (active) 45.6 (inactive) Total sample M not reported	10,308	M =26.6	Self-report: questionnaire regarding frequency and duration of PA, assigned MET weightings and used to estimate PA-related energy expenditure per week.	Classified as active if meeting WHO criteria of ≥ 2.5 hours of moderate to vigorous PA/week, an inactive if not.	Memory: 20-word recall test EF: Alice Heim 4-I test of numerical and mathematical reasoning Fluency: letter and category fluency Global cognition: Raw scores were z-transformed and averaged together for continuous measure of CF	Global CF score entered as continuous measure of CF	No association between PA and cognitive decline, or risk of dementia. However, PA levels decline in dementia cases 9 years prior to diagnosis, suggesting reverse causation.
(Tan et al., 2017)	USA	71 (male) 72 (female) Total sample M not reported	3,714	M = 7.5 years	Self-report: questionnaire regarding frequency and duration of PA, assigned MET weightings to create a PA index. Scores on this index were divided into sex specific quintiles.	PA index quintile entered as ordinal exposure.	Battery of cognitive tests and dementia clinical assessment. Incident dementia was the outcome variable.	Positive screen on test battery and diagnosis of dementia	Negative association between PA and risk of dementia. No association in sub-analysis of longer-term follow-up (>10 years). Positive associations between PA and total brain and hippocampal volume
(Thibeau, McFall, Wiebe, Anstey, & Dixon, 2016)	Australia	70.47	577	M= 4.4 years	Self-report: questionnaire regarding frequency of participation in everyday PA over a 2-year period. Total scores reflect	Total PA frequency score entered as continuous measure.	EF (Hayling and Brixton tests, Stroop test and Colour Trails test).	Latent factor of EF scores extracted and used as continuous measure of CF	Positive association between PA and EF.

Study	Country	Age at baseline	N at baseline	Follow-up time[a]	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings
					higher participation in PA.				
(Verghese, Wang, Katz, Sanders, & Lipton, 2009)	USA	78.7 (no dementia at follow-up) 79.9 (dementia at follow-up) Total sample M not reported	488	M = 4.1 years	Self-report: Self-report: questionnaire regarding frequency of participation in PA. Scores summed to create PA index.	PA index score examined continuously and dichotomised at median-split.	Battery of cognitive tests including Global cognition (Blessed Test) and dementia assessment. Incident vascular cognitive impairment or outcome variable.	Reduction in cognitive test scores, e.g. drop of >3 Blessed Test points and meeting criteria for vascular cognitive impairment.	No association between PA and risk of vascular cognitive impairment
(Willey et al., 2016)	USA	70.6	1,228	5 years	Self-report: questionnaire regarding frequency and duration of PA, assigned MET weightings. Moderate-heavy PA was activities ≥ 6 MET, light was activities <6 MET. For analytical purposes no activity and light activity were combined as the referent group.	For analytical purposes no activity and light activity were combined and used as a categorical exposure representing inactivity (the moderate-heavy group was the referent).	Cognitive battery (including Colour Trails Test, phonemic and category fluency, and Boston Naming Test) to create domain scores for: EF, PS, Episodic and semantic memory	Impairment defined as a Z score of <1.5 on any domain score.	Negative association between PA and cognitive decline.

Notes for table

[a] Follow up time refers to mean (M) or median if reported by authors. Otherwise the figure pertains to the upper range of the follow-up period.

Abbreviations: 3MS, Modified Mini Mental State Exam; 6-CIT, Six-item Cognitive Impairment Test; BSR, Buschke-Fuld Selective Remining test; CAMCOG, Cambridge Cognition Examination; CASI, Cognitive Ability Screening Instrument; CF, Cognitive function; DSST, Digit Symbol Substitution Test; EBT, East Boston Story test; EF, Executive function; M, Mean; MCI, Mild Cognitive Impairment; MET, Metabolic equivalent of task; MMSE, Mini Mental State Exam; PA, Physical activity; PACC, Preclinical Alzheimer Cognitive Composite; PS, Processing speed; SDMT, Symbol Digit Modalities Test; SD, Standard deviation; TICS, Telephone Interview for Cognitive Status; WHO, World Health Organisation

Table 2: Covariates and modelling practices of the included studies

Study	Covariates included	Modelling and adjustment strategies
(Barha et al., 2020)	Age, BMI, Cerebrovascular disease, Cardiovascular disease, Depression, Diabetes, Education	Latent growth curve modelling
(Etgen et al., 2010)	Age, Alcohol, BMI, Baseline CF, Depression, Diabetes, Heart disease, Hypertension, Hyperlipidemia, Kidney Disease, Nursing home status, Smoking, Triglycerides	Multiple regression
(Fiocco et al., 2012)	Age, BMI, Calcium, Cholesterol, Diabetes, Diastolic BP, Diet quality, Education, Energy intake (calories), Heart disease, Hypertension, Smoking, Sodium, Systolic BP	Multiple regression
(Gow et al., 2017)	Age -11 IQ, Alcohol, Education, SES, Sex, Smoking	Latent growth curve modelling
(Halloway et al., 2017)	Acculturation, Age, Chronic health problems, Depression	Multiple regression
(Hamer et al., 2018)	Age, Alcohol, BMI, Chronic lung disease, Diabetes, Depression, Education, Hypertension, Income, Sex, Smoking, Stroke, Time (follow-up)	Multiple regression
(Iso-Markku et al., 2016)	Age, BMI, Binge-drinking, Education, Hypertension, Living alone, Smoking, Time (follow-up)	Multiple regression
(Larson et al., 2006)	APOEε4 status, Age, Alcohol, Cerebrovascular disease, Depression, Diabetes, Education, Fish oil, Heart disease, Hypertension, Physical performance, Self-rated health, Smoking, Vitamin use	Multiple regression
(Morgan et al., 2012)	Age, Anxiety (state & trait), Alcohol, BMI, Cardiovascular disease, Marital status, Mental disorder, Premorbid CF, Smoking	Multiple regression
(Ogino et al., 2019)	APOEε4 status, Age, Alcohol, BMI, Depression, Education, Diabetes, Insulin medication, Heart disease, Head injury, Hypertension, Psychiatric disease, Smoking, Time (follow-up)	Multiple regression
(Papenberg et al., 2016)	Age, Cardiovascular burden, Diabetes, Education, IADL, Inflammation, MRI (grey matter volume of PFC, hippocampus, caudate and putamen), Sex	Structural Equation Modelling: PA modelled as moderator of inflammation
(Podewils et al., 2005)	ADL., IADL, APOEε4 status, Age, Baseline CF, Education, Ethnicity, MRI (white matter grade), Sex, Social network, Social support	Multiple regression
(Rabin et al., 2019)	APOEε4 status, Age, BMI, BP, Cardiovascular risk, Depression, Diabetes, Education, MRI (grey matter volume, PET (amyloid beta burden)	Multiple regression: PA modelled as moderator of AB burden risk
(Rajan et al., 2015)	ADL, Age, Cardiovascular risk, Cognitive activity, Incapacity, Physical function, Race, SES, Sex	Multiple regression
(Reas et al., 2019)	Age, Alcohol, BMI, Cardiovascular risk, Education, Sex, Smoking,	Multiple regression
(Rovio et al., 2005)	APOEε4 status, Age, Alcohol, BP, Cholesterol, Dementia, Diabetes, Education, Heart attack, Locomotor disorder, Smoking, Stroke, Time (follow-up)	Multiple regression

Study	Covariates included	Modelling and adjustment strategies
(Sabia et al., 2017)	Age, Alcohol, Antihypertensive medication, BP, Cardiovascular disease, Diabetes, Diet, Education, Ethnicity, Marital status, SES, Self-rated physical function, Smoking, Stroke	Multiple regression
(Tan et al., 2017)	APOEε4 status, Age, Cardiovascular disease, Diabetes, Education, MRI (hippocampal volume), Plasma homocysteine, Stroke	Multiple regression
(Thibeau et al., 2016)	Age, Education, Genetics (IDE & BDNF genes)	Latent growth curve modelling and multiple regression
(Verghese et al., 2009)	Age, Baseline CF, Cognitive Activity, Education, Medical illnesses, Sex	Multiple regression
(Willey et al., 2016)	Age, Alcohol, BMI, Crystallised ability, Education, Hypertension, MRI (atrophy, infarct, white matter disease)	Multiple regression

Notes for table

Abbreviations: APOE, Apolipoprotein E; BDNF, Brain-derived neurotrophic factor; BMI, Body Mass Index; BP, Blood pressure; CF, Cognitive function; IADL, Instrumental activities of daily living; IDE, Insulin degrading enzyme; IQ, Intelligence Quotient; MRI, Magnetic Resonance Imaging; PA, Physical activity; PET, Positron Emission Tomography; SES, Socio-economic status
APOE, Apolipoprotein E

Medical conditions

The most common medical variables included related to cardiovascular and cerebrovascular risk, in line with theoretical mechanism of PA's protective effect. Studies frequently used self-report or medical records to establish diagnoses of, for example, heart disease, hypertension and stroke. Biomarkers were also used to capture cardiovascular risk including blood-pressure, cholesterol and calcium. Cardiometabolic conditions such as diabetes and chronic kidney disease were also frequently accounted for. Other factors less frequently included were head injury, inflammation, locomotor disorders and musculoskeletal conditions.

Mental health and psychiatric conditions

Depression was included by seven studies. One study included anxiety, and one reported psychiatric disease as a covariate without specifying which conditions were included in this category.

Brain health

Five studies included MRI measures of brain health. These mainly related to the volume of grey matter, including at the whole-brain level and in regions of specific interest such as the

hippocampus, pre-frontal cortex, caudate and putamen. Two studies also included MRI measures of white-matter integrity, such as the volume of hyperintensities. Other MRI measures included number of silent infarcts identified. One study included a measure of β-Amyloid burden, using Positron Emission Tomography imaging.

Earlier life variables

Several studies took account of the contribution of engagement in PA at age periods earlier than baseline (assessed using retrospective questionnaires), including during teenage years, at age 30 and in middle-age. One study included childhood IQ as a measure of early-life CF, and another specifically included a measure of crystallised abilities to represent CF across the lifespan.

Other variables added to the model by consensus

Based on existing knowledge of factors associated with CF, the following variables were identified in addition to those which emerged from the review process: family history of dementia, pollution (Power, Adar, Yanosky, & Weuve, 2016), maternal smoking (Corrêa et al., 2021), childhood trauma (Cassiers et al., 2018) and psychotropic medication (Cullen et al., 2015). A generic conceptual variable, 'ancestry', was also depicted within the model to represent various historical factors influencing ethnicity, genome and familial history. Additionally, as per the translation phase of the ESC-DAGs protocol, certain conceptually similar variables were recombined into single nodes. For example, BMI and waist circumference were combined into a single node 'adiposity' and energy intake, triglycerides, and vitamin use were combined into the node 'diet'. Making recombination decisions at the conceptual DAG stage did not preclude entering each component of a recombined node separately at the analysis stage (if there were matches for these within the available Biobank data).

Synthesis of findings to produce a conceptual model

Synthesis was performed according to the ESC-DAGs method (demonstrated in figure 2. In short, this required mapping the conclusions of the 21 included studies in the form of individual DAGs, extracting the total range of covariates into an index, and assessing the relationships between all possible pairs of nodes against a set of causal criteria. The

decisions were then depicted in an integrated conceptual DAG, presented in figure 5, below. It is worth clarifying some of the assumptions that underpinned the decisions within the model. A life-course perspective was taken to group variables together according to the life-period during which they are assumed to occur, such that temporality flows left to right within the diagram (Tennant et al., 2019). Each life-period was assigned a numeric prefix for pragmatic reasons in organising the decision index and resulting code for the model. These were as follows:

- '00_': Factors which occurred prior to birth (e.g., assignment of sex at conception).
- '01_': Factors which occurred in early life (e.g., education, traumatic events, and development of traits).
- '02_': Sociodemographic factors which are realised into adulthood (e.g., socioeconomic status, employment, and marital status).
- '03_': Early adult (approximately 18-40) health behaviours which are closely associated with sociodemographic factors (e.g., PA participation, alcohol consumption and smoking status).
- '04_': Adult health outcomes which are closely associated with earlier health behaviours (and earlier factors).
- '05_': Relevant medications which are naturally associated with health status (e.g., psychotropic and antihypertensive medication).
- '06'_ & '07': The exposure period, representing baseline (middle-aged) level of PA and CF as well as intermediates between these two nodes (brain health, anxiety, sleep and mood).

The assumptions above are illustrated in DAG format in figure 4, below. The full conceptual model is presented in figure 5.

Figure (box) 4*:* Temporal assumptions within the model

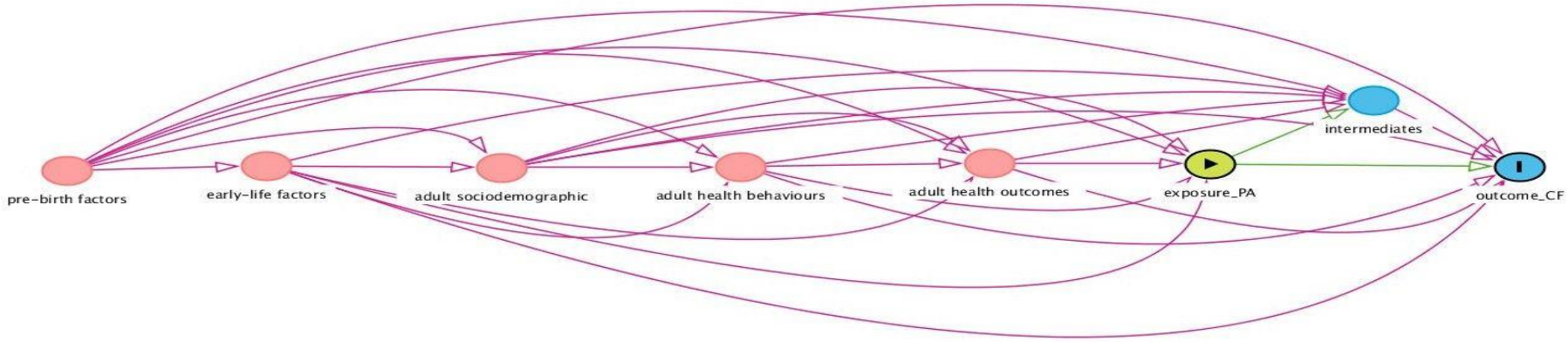


Figure 4.1: DAG representing the life-course temporal assumptions of the model

Factors grouped within each time point in time can affect those later in time either directly, or indirectly through intermediates: the past affects the future, the future does not affect the past. Note, that the minimal adjustment set is also a function of what is 'set' as the effect of interest (according to the research question). Figure 4.2, below, encodes exactly the same assumptions regarding structure of causation, yet by 'setting' the exposure earlier in time, a larger proportion of the graph is downstream, and thus a part of the effect of interest.

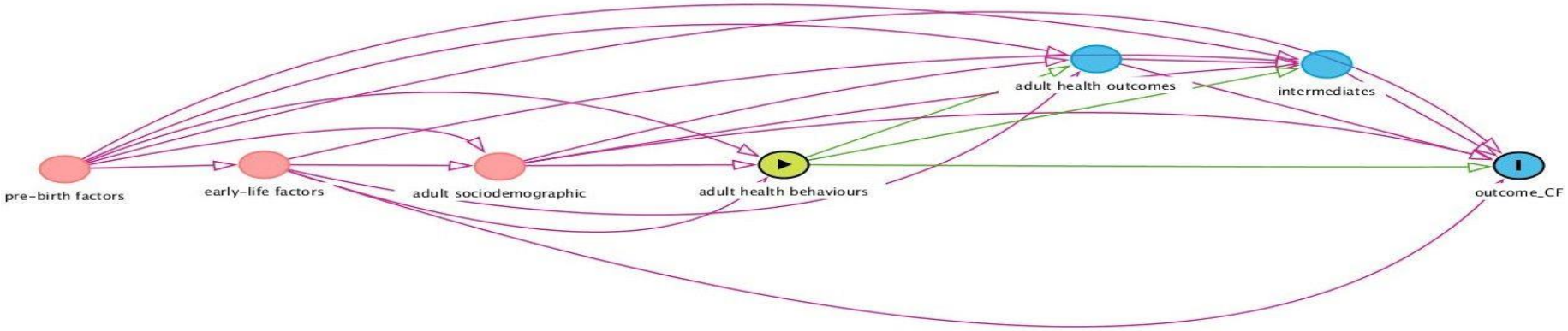


Figure 4.2: DAG representing the implications of setting a different exposure

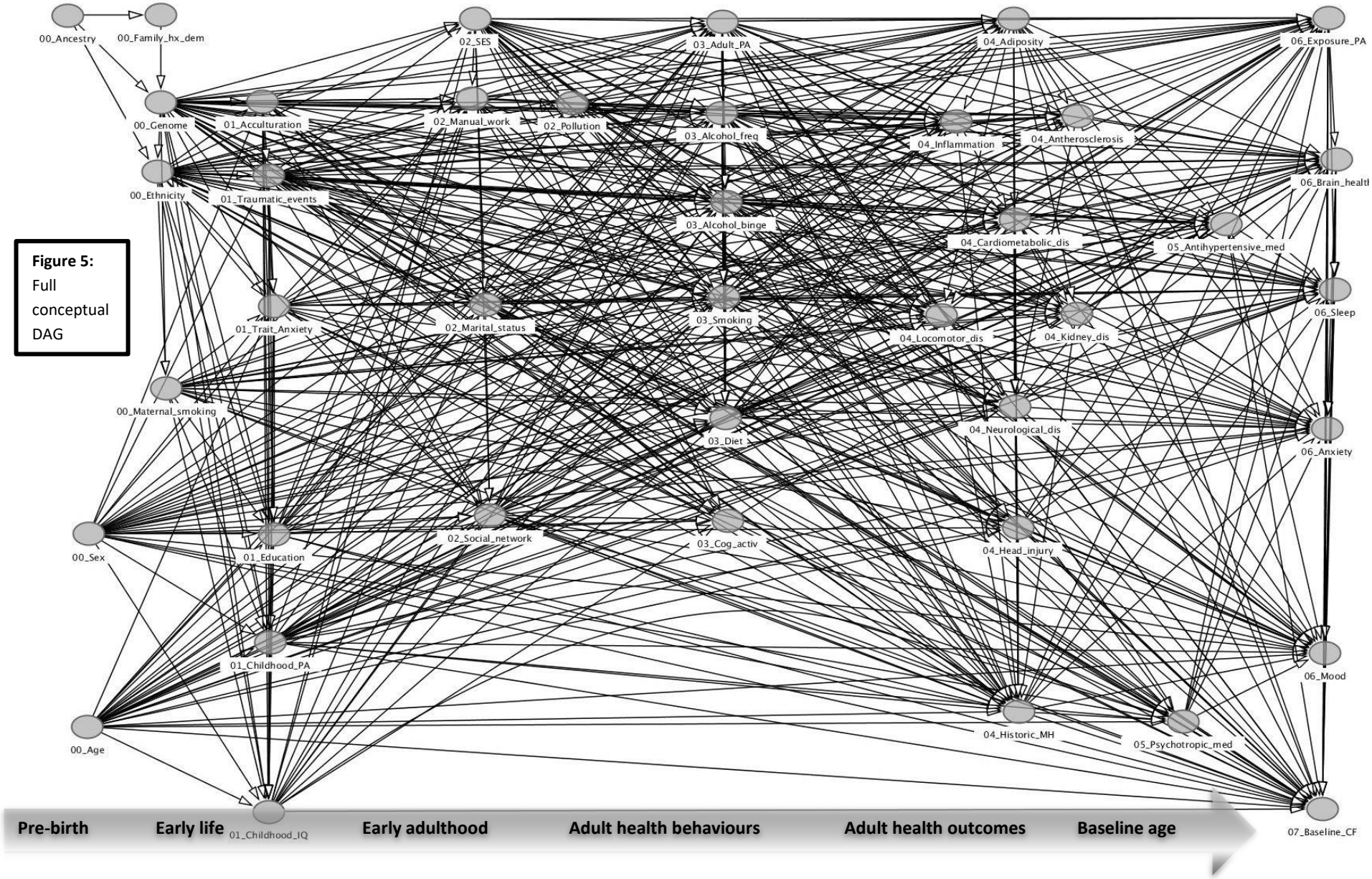


Figure 5: Full conceptual DAG

Discussion

Main findings

The primary aims of this review were to elucidate the factors associated with PA and CF in longitudinal studies of middle-aged to older adults and synthesise these findings to produce a structural causal model, following a recently developed best-practice method to do so (Ferguson et al., 2020). In the 21 included studies, factors associated with PA and CF included a broad range of biopsychosocial variables, that spanned the entire lifespan. Commonly included biological factors included genetic variants associated with CF, measures of brain health and cardiometabolic risk factors such as stroke, heart disease and diabetes. Psychological variables included mental health, particularly depression. Sociodemographic variables commonly included were age, education and ethnicity, whilst health behaviours such as alcohol consumption and smoking status were often included. Studies typically used multiple regression to adjust for these factors and estimate the association between PA and CF. No study employed a causal inference approach such as using DAGs to guide their adjustment strategies. Authors often suggested support for hypothetical mediators of the PA effect when regression coefficients were smaller after adjusting for potential mediators (e.g., cardiovascular risk factors). However, no study formally tested such hypotheses by performing mediation analysis. The majority of studies reported some form of protective association between PA and CF. Two of those which reported no association benefited from reduced risk of reverse causation due to their younger baseline samples and longer follow-up periods.

Summarising the model

Given the complexity of the conceptual model it is beyond the scope of this thesis to discuss all the decisions made during the synthesis process. However, key assumptions regarding the structure of the model are discussed below. The model leveraged the causal criterion of temporal validity to simplify potential bidirectionality between concepts and achieve acyclicity. However, it is acknowledged that there remain multiple instances of plausible bidirectional relationships within the model (e.g., the posited direction of marital status-social network could plausibly be reversed). Having established assumptions regarding temporal order, it was further assumed that each variable affected those downstream of it

directly, unless its effect was entirely captured by an intermediate. For instance, the absence of direct arrows between acculturation (early life) and various adult health outcomes encodes the assumption that the effects of acculturation on health outcomes are entirely captured by sociodemographic and health behaviour intermediate variables.

Overall, the model reflects existing literature on theoretical pathways between PA and CF such as via reduction of cardiovascular risk (Yaffe et al., 2014), improved brain health (Erickson et al., 2019) and psychological factors such as mental health and sleep (Stillman et al., 2016) and, also, the possibility that CF may be differentially sensitive to PA participation at different life-periods (Gow et al., 2017). This latter point means that decisions regarding adjustment are necessarily a function of which life period is designated as the exposure of interest (as discussed in figure 4). For instance, according to the model, if one aims to estimate the effect of early adulthood PA on mid-life to early old-age CF, then the health outcomes are downstream of PA and thus are mediators. Whereas, to estimate the effect of PA occurring more proximally to baseline CF, these health outcomes become confounders (reflecting the assumption that these medical conditions reduce PA participation and harm CF). In this thesis the decision regarding which period of PA participation is designated as the exposure must be made in the context of the empirical data used in chapter two.

Strengths, limitations, and recommendations

The use of a causal inference approach to explore the effect of PA on CF represents a novel contribution to the literature. The construction of the DAG was conducted using a systematic and transparent method that addresses the lack of guidance which has been identified in existing literature (Tennant et al., 2019). However, a consequence of this level of rigour was a highly complex model which posed practical limitations in terms of detailed interrogation of different versions of the model (due to DAGitty software crashing). This meant that the implied adjustment sets that would follow from alternative assumptions regarding the direction of causation between pairs of variables that were plausibly bi-directional (i.e., the arrow between them could have been reversed), were not fully interrogated. Based on this experience of applying the methodology, it is the opinion of the first author that the full ESC-DAGs protocol, whilst certainly robust, would not be feasible to use routinely. In particular, researchers considering using the method should be aware that

if the requirement to begin DAG construction from the assumption of full saturation is observed, then models above a certain number of nodes (approximately 30) become unfeasible to interrogate using the browser version of DAGitty. It would be helpful to consider if a 'middle ground' could be reached, i.e., a version of ESC-DAGs that balances the rigour of the protocol with pragmatism required for wider use. It should also be noted that, despite being subject to a set of rigorous and transparent criteria, the assessment and decisions of relationships between nodes are made by one researcher only. Therefore, in future applications of the method it may be of interest to conduct an inter-rater reliability check on some, or all, of the decisions made.

There were further methodological limitations to this review. The way eligible studies were prioritised for synthesis according to how many covariates they contained represented a pragmatic way of capturing the range of relevant covariates with the fewest articles but meant that articles of potentially higher methodological quality were excluded. Finally, the screening process, ROB rating and data accuracy check was only completed for a percentage of the total articles, the search strategy did not utilise forward citation searching (meaning some eligible articles may have been missed) and numbers pertaining to each reason that studies were excluded at full stage were not recorded (a deviation from PRISMA guidance).

Next steps

In chapter two of this thesis the conceptual model is taken forward and used to inform an empirical analysis using UK Biobank data, which represents the integration phase of the ESC-DAGs method.

Conclusions

This review captured a broad array of covariates that are relevant in modelling the effect of PA on CF, and transparently represented assumptions regarding the causal structure of these variables in a graphical model. Methodological limitations mean some relevant factors may not have been captured. The resulting model is taken forward to inform empirical analysis in chapter two of this thesis.

References

Barha, C. K., Best, J. R., Rosano, C., Yaffe, K., Catov, J. M., & Liu-Ambrose, T. (2020). Sex-specific relationship between long-term maintenance of physical activity and cognition in the health ABC study: potential role of hippocampal and dorsolateral prefrontal cortex volume. *The Journals of Gerontology: Series A, 75*(4), 764-770.

Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: systematic review. *Annals of epidemiology, 17*(4), 278-288.

Blondell, S. J., Hammersley-Mather, R., & Veerman, J. L. (2014). Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. *BMC public health, 14*(1), 1-12.

Cassiers, L. L., Sabbe, B. G., Schmaal, L., Veltman, D. J., Penninx, B. W., & Van Den Eede, F. (2018). Structural and functional brain abnormalities associated with exposure to different childhood trauma subtypes: A systematic review of neuroimaging findings. *Frontiers in psychiatry, 9*, 329.

Corrêa, M. L., Soares, P. S. M., da Silva, B. G. C., Wehrmeister, F., Horta, B. L., & Menezes, A. M. B. (2021). Maternal smoking during pregnancy and intelligence quotient in offspring: a systematic review and meta-analysis. *Neurotoxicology*.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., . . . Sallis, J. F. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise, 35*(8), 1381-1395.

Cullen, B., Nicholl, B., Mackay, D., Martin, D., Ul-Haq, Z., McIntosh, A., . . . Evans, J. (2015). Cognitive function and lifetime features of depression and bipolar disorder in a large population sample: Cross-sectional study of 143,828 UK Biobank participants. *European Psychiatry, 30*(8), 950-958.

Dekkers, O. M., Vandenbroucke, J. P., Cevallos, M., Renehan, A. G., Altman, D. G., & Egger, M. (2019). COSMOS-E: guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. *PLoS medicine, 16*(2), e1002742.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology, 40*, 31-53.

Erickson, K. I., Hillman, C., Stillman, C. M., Ballard, R. M., Bloodgood, B., Conroy, D. E., . . . Powell, K. E. (2019). Physical activity, cognition, and brain outcomes: a review of the 2018 physical activity guidelines. *Medicine and science in sports and exercise, 51*(6), 1242.

Etgen, T., Sander, D., Huntgeburth, U., Poppert, H., Förstl, H., & Bickel, H. (2010). Physical activity and incident cognitive impairment in elderly persons: the INVADE study. *Archives of internal medicine, 170*(2), 186-193.

Ferguson, K. D., McCann, M., Katikireddi, S. V., Thomson, H., Green, M. J., Smith, D. J., & Lewsey, J. D. (2020). Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *International journal of epidemiology, 49*(1), 322-329.

Fiocco, A. J., Shatenstein, B., Ferland, G., Payette, H., Belleville, S., Kergoat, M.-J., . . . Greenwood, C. E. (2012). Sodium intake and physical activity impact cognitive maintenance in older adults: the NuAge Study. *Neurobiology of aging, 33*(4), 829. e821-829. e828.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., . . . Chertkow, H. (2006). Mild cognitive impairment. *The Lancet, 367*(9518), 1262-1270.

Gow, A. J., Pattie, A., & Deary, I. J. (2017). Lifecourse activity participation from early, mid, and later adulthood as determinants of cognitive aging: The Lothian Birth Cohort 1921. *The Journals of Gerontology: Series B, 72*(1), 25-37.

Halloway, S., Wilbur, J., Schoeny, M. E., & Barnes, L. L. (2017). The relation between physical activity and cognitive change in older Latinos. *Biological research for nursing, 19*(5), 538-548.

Hamer, M., Terrera, G. M., & Demakakos, P. (2018). Physical activity and trajectories in cognitive function: English longitudinal study of ageing. *J Epidemiol Community Health, 72*(6), 477-483.
Harada, C. N., Love, M. C. N., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in geriatric medicine, 29*(4), 737-752.
Hernán, M. A. (2018). The C-word: scientific euphemisms do not improve causal inference from observational data. *American journal of public health, 108*(5), 616-619.
Husain, M., & Schott, J. M. (2016). *Oxford textbook of cognitive neurology and dementia*: Oxford University Press.
Iso-Markku, P., Waller, K., Vuoksimaa, E., Heikkilä, K., Rinne, J., Kaprio, J., & Kujala, U. M. (2016). Midlife physical activity and cognition later in life: a prospective twin study. *Journal of Alzheimer's Disease, 54*(4), 1303-1317.
Larson, E. B., Wang, L., Bowen, J. D., McCormick, W. C., Teri, L., Crane, P., & Kukull, W. (2006). Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older. *Annals of internal medicine, 144*(2), 73-81.
Lista, I., & Sorrentino, G. (2010). Biological mechanisms of physical activity in preventing cognitive decline. *Cellular and molecular neurobiology, 30*(4), 493-503.
Liu, C.-C., Kanekiyo, T., Xu, H., & Bu, G. (2013). Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature Reviews Neurology, 9*(2), 106-118.
Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., . . . Lisy, K. (2017). Chapter 7: Systematic reviews of etiology and risk. *Joanna Briggs Institute Reviewer's Manual. The Joanna Briggs Institute, 5*.
Morgan, G. S., Gallacher, J., Bayer, A., Fish, M., Ebrahim, S., & Ben-Shlomo, Y. (2012). Physical activity in middle-age and dementia in later life: findings from a prospective cohort of men in Caerphilly, South Wales and a meta-analysis. *Journal of Alzheimer's Disease, 31*(3), 569-580.
Nichols, E., Szoeke, C. E., Vollset, S. E., Abbasi, N., Abd-Allah, F., Abdela, J., . . . Asgedom, S. W. (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology, 18*(1), 88-106.
Northey, J. M., Cherbuin, N., Pumpa, K. L., Smee, D. J., & Rattray, B. (2018). Exercise interventions for cognitive function in adults older than 50: a systematic review with meta-analysis. *British journal of sports medicine, 52*(3), 154-160.
Ogino, E., Manly, J. J., Schupf, N., Mayeux, R., & Gu, Y. (2019). Current and past leisure time physical activity in relation to risk of Alzheimer's disease in older adults. *Alzheimer's & Dementia, 15*(12), 1603-1611.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj, 372*.
Papenberg, G., Ferencz, B., Mangialasche, F., Mecocci, P., Cecchetti, R., Kalpouzos, G., . . . Bäckman, L. (2016). Physical activity and inflammation: effects on gray‐matter volume and cognitive decline in aging. *Human brain mapping, 37*(10), 3462-3473.
Pearl, J. (2009). *Causality*: Cambridge university press.
Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*: John Wiley & Sons.
Podewils, L. J., Guallar, E., Kuller, L. H., Fried, L. P., Lopez, O. L., Carlson, M., & Lyketsos, C. G. (2005). Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. *American journal of epidemiology, 161*(7), 639-651.
Power, M. C., Adar, S. D., Yanosky, J. D., & Weuve, J. (2016). Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: a systematic review of epidemiologic research. *Neurotoxicology, 56*, 235-253.

Critical Appraisal Skills Programme (2018). CASP cohort study checklist. In: Critical Appraisal Skills Programme Oxford, UK. Available at : https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf , accessed 25/07/2021

Rabin, J. S., Klein, H., Kirn, D. R., Schultz, A. P., Yang, H.-S., Hampton, O., . . . Hedden, T. (2019). Associations of physical activity and β-amyloid with longitudinal cognition and neurodegeneration in clinically normal older adults. *JAMA neurology, 76*(10), 1203-1210.

Rajan, K. B., Barnes, L. L., Skarupski, K. A., De Leon, C. F. M., Wilson, R. S., & Evans, D. A. (2015). Physical and cognitive activities as deterrents of cognitive decline in a biracial population sample. *The American Journal of Geriatric Psychiatry, 23*(12), 1225-1233.

Reas, E. T., Laughlin, G. A., Bergstrom, J., Kritz-Silverstein, D., & McEvoy, L. K. (2019). Physical activity and trajectories of cognitive change in community-dwelling older adults: The Rancho Bernardo Study. *Journal of Alzheimer's Disease, 71*(1), 109-118.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27-42.

Rovio, S., Kåreholt, I., Helkala, E.-L., Viitanen, M., Winblad, B., Tuomilehto, J., . . . Kivipelto, M. (2005). Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. *The Lancet Neurology, 4*(11), 705-711.

Sabia, S., Dugravot, A., Dartigues, J.-F., Abell, J., Elbaz, A., Kivimäki, M., & Singh-Manoux, A. (2017). Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *Bmj, 357*.

Sexton, C. E., Betts, J. F., Demnitz, N., Dawes, H., Ebmeier, K. P., & Johansen-Berg, H. (2016). A systematic review of MRI studies examining the relationship between physical fitness and activity and the white matter of the ageing brain. *Neuroimage, 131*, 81-90.

Sofi, F., Valecchi, D., Bacci, D., Abbate, R., Gensini, G. F., Casini, A., & Macchi, C. (2011). Physical activity and risk of cognitive decline: a meta‐analysis of prospective studies. *Journal of internal medicine, 269*(1), 107-117.

Stillman, C. M., Cohen, J., Lehman, M. E., & Erickson, K. I. (2016). Mediators of physical activity on neurocognitive function: a review at multiple levels of analysis. *Frontiers in human neuroscience, 10*, 626.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *Plos med, 12*(3), e1001779.

Tan, Z. S., Spartano, N. L., Beiser, A. S., DeCarli, C., Auerbach, S. H., Vasan, R. S., & Seshadri, S. (2017). Physical activity, brain volume, and dementia risk: the Framingham study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 72*(6), 789-795.

Tennant, P. W., Harrison, W. J., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., . . . Textor, J. (2019). Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. *medRxiv*.

Textor, J., Hardt, J., & Knüppel, S. (2011). DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology, 22*(5), 745.

Thibeau, S., McFall, G. P., Wiebe, S. A., Anstey, K. J., & Dixon, R. A. (2016). Genetic factors moderate everyday physical activity effects on executive functions in aging: Evidence from the Victoria Longitudinal Study. *Neuropsychology, 30*(1), 6.

Verghese, J., Wang, C., Katz, M. J., Sanders, A., & Lipton, R. B. (2009). Leisure activities and risk of vascular cognitive impairment in older adults. *Journal of geriatric psychiatry and neurology, 22*(2), 110-118.

Willey, J. Z., Gardener, H., Caunca, M. R., Moon, Y. P., Dong, C., Cheung, Y. K., . . . Wright, C. B. (2016). Leisure-time physical activity associates with cognitive decline: The Northern Manhattan Study. *Neurology, 86*(20), 1897-1903.

Yaffe, K., Vittinghoff, E., Pletcher, M. J., Hoang, T. D., Launer, L. J., Whitmer, R. A., . . . Sidney, S. (2014). Early adult to midlife cardiovascular risk factors and cognitive function. *Circulation, 129*(15), 15

Chapter 2: Major Research project

Estimating the effect of physical activity on cognitive function within UK Biobank

Prepared in accordance with the author requirements for Journal of Neuropsychology [Appendix 1.1, p.91]

Plain English summary of Major research Project

Title: Estimating the effect of physical activity on cognitive function within UK Biobank

Background:

Cognitive function (CF) describes the range of mental abilities like memory, language, problem solving and visual ability that make it possible for humans to learn and use knowledge and skills. When CF is reduced people can find it hard to participate in their daily activities and their quality of life may suffer. Because of this, researchers are interested in identifying ways of protecting CF. Existing studies have shown that people who do more physical activity (PA; e.g., exercise, housework, DIY) have better CF than those who are inactive, which might be because PA helps keep the brain healthy. However, there are lots of other factors that are related both to PA and CF, and the way that previous studies have been designed means the strength of the true relationship between PA and CF remains unclear.

The present study used a type of diagram known as a directed acyclic graph (DAG) to help identify which of these other factors need to be adjusted for when analysing data on PA and CF. Data from a very large cohort study (UK Biobank) was then used to estimate the strength of the effect of PA on CF in middle-aged and older adults.

Aims and questions:

This study aimed to estimate how much PA affected CF for people in the UK Biobank sample, firstly using data collected at the beginning of the study, and then using follow-up data collected several years later.

If it looked like there was a strong effect of PA on CF, then a secondary aim was to check how much of this was explained by structural differences in people's brains.

Methods

People reported how much PA they did, and some of these people also had their level of PA measured directly using a device worn on their wrist. People completed different mental tasks known as cognitive tests, as a measure of their CF. Some people also had their brains scanned to measure how healthy their brains were. Statistical

models were then used to see how much PA affected CF, taking account of other aspects such as age and lifestyle.

Main findings

Surprisingly, there was very little relationship between PA and CF in our study. It might be that UK Biobank participants are healthier than other groups where this effect has been observed before, or that this sample is still too young for changes in CF related to PA to be detected reliably.

Conclusions

The findings of this study may have arisen because people in UK Biobank are not representative of the wider population. This would mean that caution is required when interpreting studies using this sample. It may also be the case that previous studies over-estimated how much PA can affect CF. One way to find out could involve repeating this study as the people in UK Biobank get older to see if the expected effect emerges later.

Word count: 478 words

Estimating the effect of physical activity on cognitive function within UK Biobank

Abstract:

Physical activity (PA) has been associated with benefits for cognitive function (CF), but previous estimates of the strength of this relationship may have been biased due to limitations in modelling practices that are common amongst observational studies. The present study aimed to address this by using a rigorously constructed conceptual causal model to guide an empirical analysis estimating the effect of PA on CF in the UK Biobank cohort of middle-aged and older adults. It was hypothesised that higher PA would be associated with better CF, and that this effect would be mediated by structural differences in brain health. PA was measured subjectively by self-report and objectively using accelerometry, and CF was measured using objective cognitive tests which have been validated against widely used standardized measures. Composite CF measures were derived to represent general and domain-specific performance. The wide range of data within UK Biobank allowed a close approximation of the covariate adjustment set specified by the model to be obtained, as well as MRI measures of brain health as potential mediators. Effect coefficients were estimated using regression models (cross-sectional: n = 31,854 to 305,294 unadjusted, n = 2,548 to 29,810 adjusted; longitudinal: n = 21,225 to 30,330 unadjusted, n = 4,805 to 6,840 adjusted). Results indicated very small effect sizes of inconsistent direction. As the hypothesized effect of PA on CF was not observed consistently, mediation analysis was not conducted. Reasons for the unexpected findings are discussed in the context of previous literature, and selection bias within UK Biobank.

Word count: 250

Key words: Physical activity; cognitive function; directed acyclic graph; healthy ageing; UK Biobank

Introduction

Conceptualisation and measurement of cognitive function

The term cognitive function (CF) describes the set of mental abilities that enable the acquisition and use of knowledge and skills throughout life. The study of CF has a long history within psychology and is often described within the literature as intelligence or cognitive ability (Deary, 2001). Whichever term is favoured, it is recognised that the structure of CF is multidimensional and consists of abilities within subdomains such as memory, speed of processing, verbal ability and reasoning (Deary, 2020). The psychometric approach to cognitive testing involves measuring people's performance on various tasks that tax abilities in a particular domain (although all tests draw on multiple domains to some degree). It has been consistently found that people's performance on different tasks is positively correlated, and that the magnitude of this relationship is stronger for tests that primarily tax the same domain. This commonality across tests and domains has often been treated as a single latent factor – '*g',* which accounts for around 50% of the variance in test performance (Fawns-Ritchie & Deary, 2020). Prominent theoretical accounts of CF (Carroll, 1993), and widely used cognitive tests (Wechsler, 2008), reflect the hierarchical structure of CF in three-stratum models comprising a general factor, underpinned by broad subdomains and narrow individual test abilities. However, the level at which CF is conceptualised and measured varies according to the aims of the researchers and clinicians who use the construct. General population research is commonly concerned with the underlying general factor, whereas neuropsychologists working with clinical populations are often concerned with domain specific deficits (Vakil, 2012). Researchers in the field of cognitive aging often use screening measures to identify impairment at the global level and thus use total scores across domains (Folstein, Folstein, & McHugh, 1975), or create composite scores which take an average across domains (Halloway, Wilbur, Schoeny, & Barnes, 2017), as described in chapter one of this thesis.

Neuroanatomical correlates of CF

Whether conceptualising CF at the global level or the domain level, it is clear that these abilities are neurologically instantiated. Brain imaging studies, typically using magnetic resonance imaging (MRI), show associations between various characteristics

of the brain and CF. In terms of size, both whole-brain volume and the volume of specific regions have shown positive associations with CF. Estimates for the correlation between whole-brain volume and CF from meta-analyses range from *r* = 0.24 (Pietschnig, Penke, Wicherts, Zeiler, & Voracek, 2015) to *r* = 0.39 (Gignac & Bates, 2017). The correlation between whole-brain volume and CF in the UK Biobank general population cohort has been estimated at *r* = 0.28 (Cox, Ritchie, Fawns-Ritchie, Tucker-Drob, & Deary, 2019). Specific regions implicated in CF include the hippocampus, thalamus, caudate, putamen, pallidum and amygdala (Miller et.al., 2016). Measures of white matter integrity are also associated with CF, including tract integrity (Penke et al., 2012) and white matter damage identified by MRI hyperintensities (Puzo et al., 2019; Ritchie, Bastin, et al., 2015). A study which combined different neuroimaging measures within multivariate models found that, overall, 18-21% of variance in CF was accounted for, with brain volume contributing 12%, cortical thickness 5% and white matter hyperintensities 2% (Ritchie, Booth, et al., 2015).

Conceptualisation and measurement of physical activity

It is important to clarify the distinction between physical activity (PA) and exercise. PA is defined as "*... any bodily movement produced by skeletal muscles that requires energy expenditure ",* whereas exercise is "*physical activity that is planned, structured and repetitive*"(Casperson et. al., 1985, p.126). PA is thus a broader category which contains everyday activities such as DIY or shopping, as well as exercise. Physical inactivity is a leading risk factor for worldwide mortality and is associated with health outcomes such as obesity, diabetes, cancer and heart disease (Lear et al., 2017). Researchers typically conceptualise PA along a continuum, and tools such as the International Physical Activity Questionnaire (IPAQ) categorise activities according to their intensity as 'light', 'moderate', and 'vigorous' (Norton, Norton, & Sadgrove, 2010). The World Health Organisation (WHO) recommends that adults should undertake 150-300 minutes of moderate intensity, or 75-150 minutes of vigorous physical activity, per week (Bull et al., 2020).

Physical activity and cognitive function

There is a broad epidemiological literature examining the effects of PA on CF. A 2009 review (Hamer & Chida, 2009) included sixteen studies measuring associations between PA and neurodegenerative disease in a meta-analysis and reported significantly reduced risk in highest vs lowest activity categories for dementia, Alzheimer's disease (AD) and Parkinson's disease. In a similar study cognitive decline (rather than neurodegenerative disease diagnosis) was used as the outcome and it was found that the people who performed a high level of PA were 38% less likely to experience cognitive decline than those categorised as sedentary, and even those who performed low-to-moderate PA were 35% less likely (Sofi et al., 2011). A meta-analysis for both cognitive decline and dementia found a significant protective effect of PA for both, though this was stronger for cognitive decline (Blondell, Hammersley-Mather, & Veerman, 2014). Another study analysed the effect of PA separately for cognitive decline, AD, vascular dementia and all-cause dementia and found significant protective effects of PA, in order of magnitude, for AD, all-cause dementia and cognitive decline, and a non-significant effect for vascular dementia (Guure, Ibrahim, Adam, & Said, 2017). Most recently an umbrella review examined 24 systematic reviews and meta-analyses of longitudinal evidence of PA's effect on health outcomes including cognitive decline (6 studies), dementia (5 studies) and AD (5 studies) amongst older adults (Cunningham, O'Sullivan, Caserotti, & Tully, 2020). The authors concluded that there is convergent evidence of reduced risk of these outcomes associated with meeting the WHO guidance for PA. Chapter one of this thesis reviewed longitudinal studies of PA and CF, and found that 18 out of the 21 included papers reported some form of protective association.

Potential mechanisms

There are several pathways by which the benefit of PA on CF may operate. As already outlined, variance in CF is associated with neuroanatomical differences. One review synthesised findings of structural brain changes associated with PA and concluded that there is evidence of modification by PA in up to 80% of grey matter (Batouli & Saba, 2017). Specific regions which appear to be implicated are the hippocampus, prefrontal

cortex and caudate nucleus (Erickson, Hillman, & Kramer, 2015). A review focusing specifically on hippocampal changes found PA conferred protection against age-associated hippocampal atrophy, and that the effect was stronger in the left hippocampus (Firth et al., 2018). The conceptual model constructed in chapter one of this thesis also reflected potential intermediate pathways between PA undertaken earlier in adulthood and CF via reduction of cardiovascular risk, and PA occurring more proximally to CF measurement via psychological variables such as mood, anxiety and sleep.

Other covariates

Other factors which vary with PA and CF were systematically reviewed in chapter one of this thesis and synthesised to construct a conceptual model. There were a large number of confounders and mediators indicated by this process. Confounders broadly fell into the following categories: pre-birth factors (e.g., genetic risk); early life factors (e.g., childhood PA, education, traumatic events); adult sociodemographic factors (e.g., socioeconomic status, exposure to pollution); adult health behaviours (e.g., diet, alcohol consumption and smoking status); adult health outcomes (e.g., cardiovascular disease, neurological disease, mental health disorders); and medication (e.g., psychotropic or antihypertensive medication). Intermediates between baseline levels of PA and CF were brain health, sleep, anxiety and mood. Physical activity has long been associated with benefits for mental health (Ströhle, 2009), and evidence based interventions for depression such as behavioural activation can often cross over into PA either directly or indirectly (Lambert, Greaves, Farrand, Haase, & Taylor, 2017).

Rationale for the present study

As outlined in chapter one of this thesis, previous observational studies of the PA -CF relationship may be inaccurate because none have followed a comprehensive method to select covariates. The present study addressed this by using a graphical causal model to inform estimation of the total effect of PA on CF. Of 21 studies reviewed in chapter one, none performed formal mediation analysis to test potential intermediate mechanisms of the PA-CF relationship. The present study aimed to address this by considering potential mediation analysis to decompose any observed main effects into

direct and indirect effects, using the measures of brain health available within UK Biobank.

Aims

This study matched variables within the conceptual causal model reported in chapter one, with data available within the UK Biobank dataset, in order to address the following research aims:

1) To estimate the magnitude of the relationship between PA and CF in a cross-sectional analysis of baseline UK Biobank data
2) To estimate the magnitude of the relationship between PA at baseline and CF at follow-up in a longitudinal analysis
3) To estimate the extent to which significant PA-CF relationships (if any) were mediated by structural brain differences

Methods

This study is reported according to the Strengthening Reporting of Observational Studies in Epidemiology (STROBE) statement (Von Elm et al., 2014).

Participants

UK Biobank is a very large prospective cohort study of over 500,000 participants designed to examine the genetic and environmental determinants of health in middle-aged to older adults in the general population (Sudlow et al., 2015). Ethical approval was granted by the NHS Research Ethics Committee (appendix 2.1, p.107), and covers all studies relating to the resource. The present study was conducted under approved application 11332 (appendix 2.2, p.111). NHS Lanarkshire's R&D department was also notified that their employee was conducting this research, which was acknowledged by the health board (appendix 2.3, p.112).

Participants provided written informed consent. Recruitment was based on proximity to an assessment centre and being within the eligible age range of 40-69. During the baseline assessment (2006-2010) participants attended assessment centres around the UK where they completed self-report sociodemographic, health and lifestyle questionnaires and an interview with a trained staff member, as well as undergoing

physical and biological measurements and a brief computerised cognitive assessment. Subsequently, a subset of the total cohort completed accelerometry (objective physical activity measures) and neuroimaging visits (including repeat cognitive assessments). Invitations to participate in accelerometry (2013-2015) (Doherty et al., 2017) were sent to a random sample of participants with email addresses (excluding those closest to the main UK Biobank centre, due to concerns about burden on those participants). Invitations to the neuroimaging assessments (2014 onwards) were based on proximity to UK Biobank MRI scanning centres in England. Because the present study made use of genetic score data, the analysis sample was restricted to those with white British genetic ancestry (as determined centrally by UK Biobank based on a combination of self-report and genetic data) in order to reduce confounding induced by groups differing systematically both by genetic ancestry and according to phenotypic measures of interest (Turner et al., 2011). This type of restriction is standard practice and has been performed in UK Biobank studies previously (Milton et al., 2021). Similarly, the sample was restricted to unrelated people; this was done by randomly keeping one member of each related set (third degree or closer). Along with further standard exclusions based on genotyping quality control, this left a sample of 334,227 which was used for baseline analysis in the present study (see figure 1).

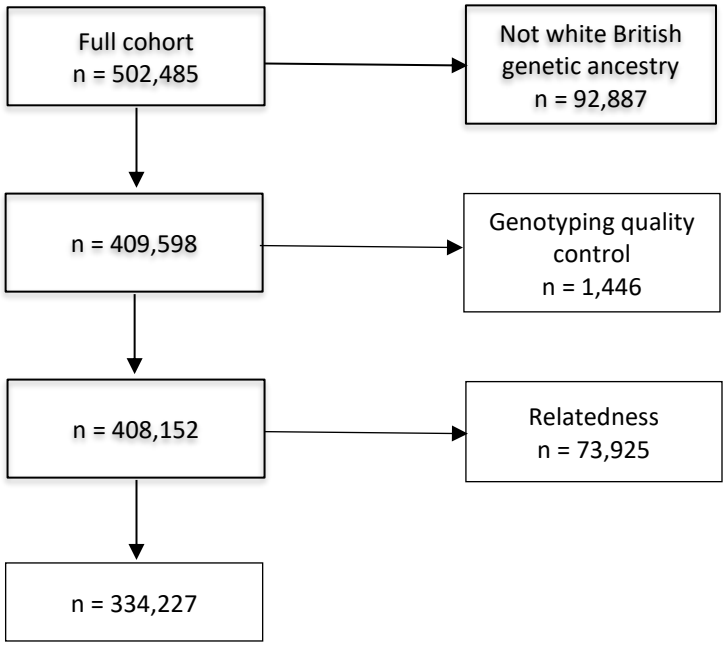


Figure 1: Flowchart showing participants included in the study

Measures

The variables within the conceptual model presented in chapter one were matched to the data available within UK Biobank (appendix 2.4, p.113).

Exposure: baseline physical activity

Self-report data: As part of the assessment centre visits, participants completed a modified form of the International Physical Activity Questionnaire (IPAQ) short form (Craig, etl.al., 2003), reporting the frequency and duration of walking, moderate and vigorous activity undertaken in a typical week. Data were processed in accordance with IPAQ scoring guidance, such that each category of activity was assigned the following weighting: walking, 3.3 METs, moderate activities, 4 METs and vigorous activities 8 METs. Total amount of moderate-vigorous PA was estimated as the sum of these moderate and vigorous PA expressed in MET hours per week, and classified as active if they met IPAQ recommendations of at least 10 MET hours per week of moderate to vigorous PA, as has been done in previous UK Biobank studies (Celis-Morales et al., 2019).

Total PA was calculated by summing the weighted time spent across all three categories and expressed in MET-hours per week. Therefore, participants receive a total PA value if they had data for at least one of the three levels of activity.

Accelerometer data: A subsample participated in accelerometer-measured PA data collection for a one-week period. The wrist-worn actigraph device recorded mean daily accelerations, expressed in milli-gravity per day, which was used as the objective measure of total PA (Doherty et al., 2017).

Outcome: cognitive function

Cognitive tests have been administered at several time-points to Biobank participants. These tests were developed specifically for Biobank, to enable computerised administration at scale without staff involvement, and are thus non-standardised.

However, the psychometric properties of these measures have since been compared to well-validated reference tests (Fawns-Ritchie & Deary, 2020). For the present study, the raw scores for all tests except prospective memory (as it is a binary variable) were converted into z-scores for ease of interpretation, standardised within five-year age bands at each assessment time-point. Therefore, the mean score is approximately zero, and the standard deviation for each is approximately one. For each z-score, higher scores represent better performance.

Baseline CF

During the original baseline assessment, almost all participants completed touchscreen tests of visuospatial memory ('Pairs Matching') and processing speed ('Reaction Time'). Additional tests of prospective memory ('Prospective Memory'), attention/working memory ('Numeric Memory'), and verbal and numerical reasoning ('Reasoning') were added to the battery part way through the recruitment period, one of which (Numeric Memory) was subsequently removed for reasons of time. The sample sizes on these latter three tests are therefore smaller.

- *Reaction Time:* Participants were shown pairs of cards on a screen and asked to press a button as quickly as possible when the two cards were identical. Twelve pairs were shown in total.
- *Pairs Matching:* Participants were shown 12 cards onscreen simultaneously and were asked to recall the position of six matching pairs.
- *Reasoning:* Participants were given two minutes to answer 13 multiple-choice verbal and numerical reasoning questions. UK Biobank refers to this as a fluid intelligence test; however some questions require crystallised abilities, and thus the task has been described as a reasoning test here, in line with other studies (Lyall et al., 2016).
- *Numeric Memory:* Participants were presented with a string of numbers onscreen, and asked to enter them on a keypad in reverse order from memory, following a brief delay. This test was intended to require participants to mentally reverse the numbers, making it similar to

backward digit span tasks. However, the stimuli were actually presented simultaneously rather than sequentially meaning the participants were able to achieve the correct response by recalling digit strings without reversing them if they read the numbers from right to left. This represents a forward digit span task which is an easier task, reflecting the attention/working memory domain.

- *Prospective Memory:* Participants were presented with on-screen text informing them that at the end of the cognitive tests they would see four coloured symbols and be asked to touch the blue square. However, the instructions went on to inform the participant that they are to touch the orange circle instead. This required the participant to recall and respond in accordance with the true objective of the test.
- *Global CF:* For the present study a composite measure of global CF was created by taking the mean of the four baseline z-scores, as has been done in previous studies (Halloway et al., 2017; Rajan et al., 2015; Sabia et al., 2017). This score was only created for participants with two or more non-missing z-score values.

Follow-up CF

The follow-up data in the present study pertains to ten tests administered at the imaging visit (the five described above plus five below).

- *Trail-making Test:* Part A required participants to click on 25 numbered circles in ascending order, reflecting processing speed. Part B involved a similar task but switching between letters and numbers, reflecting processing speed plus executive function. Scores for each part reflect time taken in seconds to correctly click all circles. Additional scores comprising Part B time minus Part A time (as a more sensitive measure of executive function) and the number of errors made on part B were also derived.
- *Digit Symbol Substitution Test:* A grid of eight symbols, each corresponding to a number, was displayed onscreen. Participants used a keypad to enter the number corresponding to each symbol as it was presented onscreen,

and scores reflect the total number of boxes correctly filled within two minutes. This task primarily reflects processing speed.

- *Tower Rearranging Test:* Participants were presented with an onscreen illustration of three pegs, upon which three coloured hoops had been placed. They were then asked how many moves would be required to rearrange the hoops into another specified configuration. This test reflects executive function.
- *Paired Associate Learning:* Participants were shown twelve pairs of words for 30 seconds in total, and after an interval presented with the first word for ten of these pairs and asked to select the matching word from lists of four alternatives. This test reflects verbal declarative memory.
- *Matrices:* Participants were presented with a series of matrix pattern blocks with an element missing and asked to select the element that best completed the pattern from a range of specified choices. This test reflects non-verbal reasoning.

One additional test (Picture Vocabulary) was administered at the imaging visits. Unfortunately, the data relating to this test have not yet been released by UK Biobank and so this is not described further here.

Global CF: A composite measure of global CF was derived by taking the mean of ten imaging visit z-scores (trails B-A not included), for participants with at least two non-missing z-score values.

Domain-level composites:

As the imaging visit data included multiple tests that measure the same domain of CF, composite scores were derived using the mean of z-scores for participants who had at least two non-missing values for tests within that domain. The following domain-level composite scores were derived: processing speed (Digit Symbol Substitution and Reaction Time); reasoning (Reasoning and Matrices); executive function (Tower Rearranging, Trails-A time and Trails-B time) and memory (Pairs Matching, Numeric Memory and Paired Associate Learning).

Intermediates

The intermediates of interest in this thesis pertained to brain health (although other pathways via mood, anxiety and sleep were also indicated by the model). All brain MRI data were acquired on a Siemens Skyra 3T scanner. The acquisition protocol is described in detail elsewhere https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/brain_mri.pdf. The variables of interest were determined by the systematic review reported in chapter one, and additional literature known to be relevant.

Grey matter volume: Measures of interest which were available within UK Biobank were total grey matter volume as well as regional volumes of the left and right hippocampus, left and right dorsolateral prefrontal cortex (represented by the superior and middle frontal gyrus) and left and right anterior cingulate cortex.

White matter: White matter measures of interest were total volume of hyperintensities, and a general factor representing tract integrity across 27 bilateral white matter tracts (created using principal components analysis).

Other covariates

The covariates were identified by the graphical causal model reported in chapter one and matched to available UK Biobank data. These are described in appendix (2.5, p.116) and listed in full in table 1.

Unmatched variables

There were several variables within the conceptual model which could not be matched to data within UK Biobank. These were childhood PA, childhood IQ, earlier adulthood PA and cognitive activity. Of these, childhood PA and earlier adulthood PA were specified in all minimum covariate sets. Thus, the model estimated in this study represents the nearest approximation of the full conceptual model, as is recommended in recent guidance (Tennant et al., 2019).

Statistical analysis

All analyses were performed in Stata version 16. Data were summarised using descriptive statistics and displayed according to PA classification: active, inactive or missing. Normally distributed continuous variables are summarised as means and standard deviations, and skewed variables as medians with inter-quartile ranges. Ordinal and binary variables are reported as frequencies and percentages within each category. These summary statistics are presented for the baseline characteristics of the total sample, and the subsample who attended the imaging visit (table 1a). Data pertaining to the cognitive outcomes at follow-up are presented in table 1b. Differences between the PA groups for each measure were not formally tested, as the decision about entering covariates into the regression models was based *a priori* on the DAG rather than on the existence of statistical differences. The relationship between PA and CF was then estimated using two sets of regression models.

The first set of regression models (table 2) used CF data that was measured cross-sectionally with the PA measure. Cognitive scores at baseline were entered as the dependent variable and total self-reported PA in MET hours per week, as a continuous independent variable. Models were initially run without adjustment, and then adjusted according to the nearest approximation of the minimum sufficient adjustment set (listed below table 2).

The second set of regression models (table 3) used the CF variables pertaining to the imaging visit, making the analysis longitudinal by design. The included covariates were as above with the addition of follow-up duration, and both self-reported PA and the covariate values were again taken from baseline data. This set of models was also repeated using accelerometery data as an objective measure of PA (which was acquired after baseline CF measurement and thus not used in cross sectional models).

Diagnostic checks were performed for all models to ensure the assumptions for regression were met. False discovery rate correction of the p values was used to minimise the rate of false positive significant results within groups of models that tested the same hypotheses, with the corrected significance level set at 0.05.

Where the above total effects models for CF at follow-up (imaging visit) showed a significant relationship between PA and CF, it was planned that mediation models would be conducted to estimate the magnitude of the effect that was transmitted via structural brain MRI measures; these models would be adjusted for confounders of the relationships between PA, CF and the mediators, as determined from the DAG.

All analyses were conducted on a complete-case basis and missing values were not imputed.

Results

Sample characteristics

Table 1a shows descriptive statistics for the variables specified in the conceptual model at baseline for both the entire sample, and for the subsample who returned for imaging. The subsample who returned for imaging were on average younger, more active, less deprived, and generally healthier at baseline than the overall sample. It is also apparent that, within the baseline data, those who were missing PA status were less educated, more deprived, and generally less healthy than the overall sample, suggesting that missingness on the moderate-vigorous PA measures (which determined the PA groups) was not random. High missingness on some cognitive tests reflects that some tests were introduced at different stages within the baseline recruitment window. Table 1b shows the cognitive outcomes for the imaging sample at follow-up. Generally, the descriptive statistics suggested very small reductions in CF, of similar magnitude across the PA groups. The mean duration between baseline and follow-up was 8.94 years (SD 1.76).

Effect of physical activity on cognitive function at baseline

Table 2 shows the cross-sectional regression results estimating the effect of PA, expressed in MET hours per week, on CF, expressed in z-score units (with the exception of Prospective Memory, which is an odds ratio reflecting the odds of a correct response). Missing values throughout the dataset resulted in different sample sizes for analysis. For baseline analyses sample sizes ranged from 31,854 (Numeric Memory) to 305,294 (Reaction Time) in unadjusted models. Adjusted model sample

sizes ranged from 2,548 (Prospective Memory) to 29,810 (Reaction Time). The unadjusted models indicated a statistically significant, but trivially small, effect of PA on CF. For each measure of CF, the direction of the effect was negative (harmful), except for Reaction Time which was positive (protective). When models were adjusted for covariates, the effects were no longer significant except for Reasoning, which remained significant and still of very small magnitude.

Effect of physical activity on cognitive function at follow-up

Table 3 displays the longitudinal regression results estimating the effect of self-reported PA, expressed in MET hours per week, on CF, expressed in z-score units (with the exception of Prospective Memory, which is an odds ratio reflecting the odds of a correct response). Results for the same models repeated using accelerometery averages (expressed in milligravity units) as a continuous measure of objectively measured PA, are displayed in the lower half of the same table.

For self-reported PA, unadjusted longitudinal model sample sizes ranged from 21,225 (Trails-B Time) to 30,330 (Prospective Memory). Across the unadjusted self-reported PA models the estimated effect of PA on CF was trivially small, albeit statistically significant (except in the case of Reaction Time which was non-significant). The direction of the effect was negative (harmful) for all CF measures. After adjusting the models for the specified covariates, sample sizes ranged from 4,805 (Paired Associate Learning) to 6,840 (Prospective Memory). Outcomes which remained significant following adjustment were Reasoning (individual tests and composite measure), Trails (B time, B-A time and B errors) and the composite Executive Function measure, and the Memory and Global composites. All of these effects were trivially small in magnitude, and in the negative direction.

For objectively measured PA, sample sizes for unadjusted models ranged from 9,362 (Trails-B errors) to 14,392 (Prospective Memory). There were trivially small but significant effects in unadjusted models for Reaction Time, Symbol Digit Substitution, Trails A time, and the Processing Speed composite, all in the positive (protective) direction. After adjusting for the specified covariates, sample sizes ranged from 2,742

(Trails-B Time) to 3,935 (Reaction Time). The effect size estimates remained trivially small, and none were statistically significant.

Mediating pathways

Given that the results from the follow-up models were inconsistent in direction and with very small effect sizes, it was decided that mediation analysis would not be conducted.

Sensitivity analysis

Because the adjusted models contained large numbers of covariates, results were potentially sensitive to bias arising from missing data. To examine this possibility a sensitivity analysis was performed by repeating the unadjusted analyses, restricted to those participants who had full covariate data. The results are presented in appendix 2.6 (p.118). There was very little difference in effect estimates, indicating that the unadjusted relationship between PA and CF was very similar among people with and without missing covariate data. Therefore, it is unlikely that observed results in adjusted models are being driven by missing data bias in the analytic sample.

Table 1a: Baseline characteristics of total sample and the imaging sub-sample

	Total sample	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**		**Total sample**	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**
	Total sample					**Imaging sub-sample (at baseline)**			
n (%) of sample)	334,227 (100.00)	181,587 (54.33)	82,917 (24.81)	69,723 (20.86)		34,058 (100.00)	19,495 (57.24)	9,175 (26.94)	5,338 (15.82)
	Demographics					**Demographics**			
Age (years) Mean (SD)	56.86 (7.99)	56.80 (8.14)	56.49 (7.84)	57.46 (7.70)		55.46 (7.50)	55.57 (7.64)	55.30 (7.28)	55.35 (7.31)
Sex n (%) female	179,421 (53.68)	94,029 (51.78)	44,798 (54.03)	40,594 (58.22)		17,317 (50.85)	9,598 (49.23)	4,682 (51.03)	3,037 (56.37)
	Physical activity					**Physical activity**			
Acceleration average (milligravity units)									
n (%) missing	262,062 (78.41)	140,156 (77.18)	63,969 (77.15)	57,937 (83.10)		18,837 (55.31)	10,606 (54.40)	5,187 (56.53)	5,388 (56.50)
Mean (SD)	27.41 (14.65)	28.75 (16.17)	25.95 (13.84)	25.09 (8.54)		28.08 (9.08)	29.26 (9.60)	26.60 (8.03)	26.08 (7.91)
mvPA, self-report (MET hrs/week)									
n (%) missing	69,723 (20.86)	0 (0)	0 (0)	69,723 (100.00)		5,388 (15.82)	0 (0)	0 (0)	5,388 (100.00)
Median	18.67	30.00	4.00	n/a		18.00	28.00	4.00	n/a
(Q1, Q3)	(8.00, 40.00)	(18.00, 56.00)	(2.00, 6.67)			(7.33, 36)	(17.33, 36.00)	(2.00, 6.67)	n/a
Total PA, self-report (MET hrs/week)									
n (%) missing	27,310 (8.17)	0 (0)	0 (0)	27,310 (39.17)		1,803 (5.29)	0 (0)	0 (0)	1,803 (33.46)
Median	28.22	47.55	11.90	8.25		27.55	43.62	11.70	7.70
(Q1, Q3)	(12.89, 58.10)	(29.30, 84.00)	(7.26, 19.00)	(4.13, 16.50)		(13.20, 53.10)	(27.95, 73.7)	(7.26, 18.50)	(3.30, 15.40)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Cognitive function (baseline tests)						**Cognitive function (baseline tests in imaging subsample)**			
Numeric memory (Z score) n (%) missing Mean (SD)	299,740 (89.68) -0.35 (0.94)	161,673 (89.03) -0.34 (0.93)	74,727 (90.12) - 0.27 (0.93)	63,340 (90.85) -0.46 (0.94)		30,634 (89.95) -0.18 (0.93)	17,460 (89.56) -0.20 (0.93)	8,283 (90.27) -0.13 (0.93)	4,892 (90.79) -0.20 (0.93)
Pairs matching (Z score) n (%) missing Mean (SD)	7,550 (2.26) 0.24 (1.03)	3,385 (1.86) 0.23 (1.04)	1,472 (1.78) 0.27 (1.03)	2,693 (3.86) 0.22 (1.04)		446 (1.31) 0.32 (1.03)	179 (0.92) 0.31 (1.03)	70 (0.76) 0.33 (1.03)	197 (3.66) 0.35 (1.01)
Prospective memory n (%) missing n (%) correct on first attempt	223,812 (66.96) 88,383 (80.05)	181,587 (65.90) 49,673 (80.23)	55,812 (67.31) 22,505 (83.03)	48,330 (69.32) 16,205 (75.75)		22,689 (66.62) 8,551 (87.68)	12,831 (65.82) 5,809 (86.73)	9,175 (66.79) 2,742 (89.75)	3,730 (69.23) 1,492 (85.21)
Reaction time (Z score) n (%) missing Mean (SD)	2,010 (0.60) 0.05 (0.95)	849 (0.47) 0.08 (0.94)	427 (0.51) 0.06 (0.94)	734 (1.05) -0.02 (0.96)		65 (0.19) 0.16 (0.93)	25 (0.13) 0.18 (0.92)	18 (0.20) 0.14 (0.92)	22 (0.41) 0.11 (0.93)
Reasoning (Z score) n (%) missing Mean (SD)	1,363 (1.24) -0.13 (0.94)	646 (1.05) -0.13 (0.93)	226 (0.84) 0.01 (0.93)	491 (2.33) -0.28 (0.94)		22,689 (66.62) 0.14 (0.89)	12,831 (65.82) 0.10 (0.90)	9,175 (66.79) 0.24 (0.87)	3,730 (69.23) 0.08 (0.90)
Global CF[b] (Z score) n (%) missing Mean (SD)	7,540 (2.26) 0.11 (0.70)	3,324 (1.83) 0.11 (0.69)	1,462 (1.76) 0.14 (0.69)	2,754 (3.95) 0.05 (0.71)		2,370 (6.96) -0.06 (0.46)	1,407 (7.22) - 0.06 (0.46)	539 (5.87) -0.03 (0.45)	424 (7.87) -0.11 (0.46)
Genetics						**Genetics**			

	Total sample	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**		**Total sample**	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**
APOE genotype n (%) missing n (%) with each number of APOE e4 alleles	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)		0 (0)	0 (0)	0 (0)	0 (0)
0	237,761 (71.14)	128,621 (70.83)	59,227 (71.43)	49,913 (71.59)		24,527 (72.02)	14,036 (72.00)	6,623 (72.19)	3,868 (71.79)
1	88,276 (26.44)	48,461 (26.69)	21,703 (26.17)	18,212 (26.12)		8,754 (25.70)	4,998 (25.64)	2,343 (25.54)	1,413 (26.22)
2	8,090 (2.42)	4,505 (2.48)	1,987 (2.40)	1,598 (2.29)		777 (2.28)	461 (2.36)	209 (2.08)	107 (1.99)
Polygenic dementia risk score (Z score) n (%) missing	1,000 (0.30)	545 (0.30)	242 (0.29)	213 (0.31)		113 (0.33)	65 (0.33)	31 (0.34)	17 (0.32)
Mean (SD)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)		-0.02 (1.00)	-0.02 (1.00)	-0.01 (1.00)	-0.01 (1.00)
Familial risk						**Familial risk**			
Dementia (parent or sibling with diagnosis) n (%) missing	48,778 (14.59)	24,599 (13.55)	10,916 (13.16)	13,263 (19.02)		3,954 (11.61)	2,198 (11.27)	1,023 (11.15)	733 (13.60)
n (%) with diagnosis	40,168 (12.02)	21,851 (12.03)	10,123 (12.21)	9,194 (11.75)		4,168 (12.24)	2,384 (12.23)	1,147 (12.50)	637 (11.82)
Maternal smoking around birth n (%) missing	46,978 (14.06)	24,009 (13.22)	11,214 (13.52)	11,755 (16.86)		4,127 (12.12)	2,281 (11.70)	1,079 (11.76)	767 (14.24)
n (%) answered yes	88,016 (26.33)	47,858 (26.36)	21,284 (25.67)	18,874 (27.07)		9,119 (26.77)	5,210 (26.72)	2,420 (26.38)	1,489 (27.64)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Parkinson's disease (parent or sibling with diagnosis)									
n (%) missing	53,562 (16.03)	26,860 (14.79)	12,045 (14.53)	14,657 (21.02)		4,205 (12.35)	2,289 (11.74)	1,081 (11.78)	835 (15.50)
n (%) with diagnosis	13,606 (4.07)	7,272 (4.00)	3,451 (4.16)	2,833 (4.13)		1,362 (4.00)	792 (4.06)	359 (3.91)	211 (3.92)
Sociodemographic						**Sociodemographic**			
Acculturation (years in UK)									
n (%) missing	0 (0)	0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)	0 (0)
Median (Q1, Q3)	58.00 (50.00, 63.00)	58.00 (50.00, 63.00)	57.00 (50.00, 63.00)	59.00 (52.00, 64.00)		56.00 (49.00, 61.00)	56.00 (49.00, 62.00)	56.00 (50.00, 61.00)	56.00 (50.00, 61.00)
Educational attainment									
n (%) missing	3,105 (0.93)	1,271 (0.70)	421 (0.51)	1,413 (2.03)		226 (0.66)	46 (0.24)	13 (0.14)	167 (3.10)
Has a degree, n (%)	106,384 (31.83)	60,045 (33.07)	30,874 (37.23)	15,465 (22.18)		15,524 (45.58)	8,991 (46.12)	4,579 (49.91)	1,954 (36.27)
Household income									
n (%) missing	46,127 (13.80)	22,274 (12.27)	8,930 (10.77)	14,923 (21.40)		2,979 (8.75)	1,566 (8.03)	628 (6.84)	785 (14.57)
n (%) in each income category									
< £18k	62,329 (18.65)	32,501 (17.90)	13,772 (16.61)	16,056 (23.03)		3,530 (10.36)	2,072 (10.63)	827 (9.01)	631 (11.71)
£18k - £30,999	73,830 (22.09)	41,779 (23.01)	17,902 (21.59)	14,150 (20.29)		6.992 (20.53)	4,083 (20.94)	1,853 (20.20)	1,056 (19.60)
£31k – £51,999	76,417 (22.86)	42,474 (23.39)	20,447 (24.66)	13,496 (19.36)		9,401 (27.60)	5,377 (27.58)	2,557 (27.87)	1,467 (27.23)
£52k - £100k	59,938 (17.93)	33,266 (18.32)	17,436 (21.03)	9,236 (13.25)		9,787 (25.80)	4,893 (25.56)	2,615 (28.50)	1,189 (22.07)
>£100k	15,586 (4.66)	9,294 (5.12)	4,430 (5.34)	1,862 (2.67)		2,369 (6.96)	1,414 (7.25)	695 (7.57)	260 (4.83)

	Total sample	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**		**Total sample**	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**
Living alone n (%) missing n (%) answered yes	981 (0.29) 60,558 (18.12)	464 (0.26) 31,583 (17.39)	190 (0.23) 15,016 (18.11)	327 (0.47) 13,959 (20.02)		51 (0.15) 5,162 (15.16)	24 (0.12) 2,904 (14.90)	11 (0.12) 1,410 (15.37)	16 (0.30) 848 (15.74)
Married n (%) missing n (%) answered yes	954 (0.29) 247,610 (74.08)	464 (0.26) 136,554 (75.20)	191 (0.23) 61,596 (74.29)	299 (0.43) 49,460 (74.08)		56 (0.16) 26,614 (78.14)	27 (0.14) 15,281 (78.38)	12 (0.13) 7,199 (78.46)	17 (0.32) 4,134 (76.74)
Pollution (inverse distance to major road) n (%) missing n (%) in each quintile (Q1 = farthest from road)	4,540 (1.36)	2,402 (1.32)	1,188 (1.43)	950 (1.36)		396 (1.16)	216 (1.11)	108 (1.18)	72 (1.34)
Qu1	70,159 (20.99)	38,825 (21.38)	17,428 (21.02)	13,906 (19.94)		6,985 (20.51)	4,057 (20.81)	1,864 (20.32)	1,064 (19.75)
Qu2	66,726 (19.96)	36,220 (19.95)	16,570 (19.98)	13,936 (19.99)		6,810 (20.00)	3,920 (20.11)	1,834 (19.99)	1,056 (19.60)
Qu3	64,998 (19.45)	35,267 (19.42)	15,923 (19.20)	13,808 (19.80)		6,423 (18.86)	3,638 (18.66)	1,730 (18.86)	1,055 (19.58)
Qu4	64,364 (19.26)	34,790 (19.16)	15,942 (19.23)	13,632 (19.55)		6,740 (19.79)	3,899 (20.00)	1,836 (20.01)	1,005 (18.65)
Qu5	63,440 (18.98)	34,083 (18.77)	15,866 (19.13)	13,491 (19.35)		6,704 (19.68)	3,765 (19.31)	1,803 (19.65)	1,136 (21.08)

	Total sample	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**		**Total sample**	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**
Social network (frequency of friend/family visits)									
n (%) missing	1,926 (0.58)	540 (0.30)	219 (0.26)	1,167 (1.67)		208 (0.61)	28 (0.14)	7 (0.08)	173 (3.21)
n (%) in each category									
Almost daily	39,054 (11.68)	21,950 (11.89)	8,534 (10.29)	8,930 (12.81)		3,244 (9.52)	1,942 (9.96)	766 (8.35)	536 (9.95)
2-4 times/week	104,315 (31.21)	58,435 (32.18)	25,277 (30.48)	20,603 (29.55)		10,162 (29.84)	6,081 (31.19)	2,650 (28.88)	1,431 (26.56)
About once a week	118,650 (35.50)	64, 863 (35.72)	30,449 (36.72)	23,338 (33.47)		12,804 (37.59)	7,270 (37.29)	3,609 (39.34)	1,925 (35.73)
About once a month	43,841 (13.12)	23,108 (12.73)	11,937 (14.40)	8,796 (12.62)		14,86 (5,060	2,794 (14.33)	1,455 (15.86)	811 (15.05)
Once every few months	21,252 (6.36)	10,717 (5.90)	5,399 (6.51)	5,136 (7.37)		2,233 (6.56)	1,219 (6.25)	607 (6.62)	407 (7.55)
Never or almost never	4,581 (1.37)	2,093 (1.15)	973 (1.17)	1,515 (2.17)		322 (0.95)	154 (0.79)	73 (0.80)	85 (1.76)
No friends/family outside household	608 (0.18)	241 (0.13)	129 (0.16)	238 (0.34)		25 (0.07)	7 (0.04)	8 (0.09)	10 (0.19)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Townsend deprivation score quintiles (Q1 = least deprived)									
n (%) missing	404 (0.12)	213 (0.12)	110 (0.13)	81 (0.12)		30 (0.09)	18 (0.09)	11 (0.12)	1 (0.02)
n (%) in each quintile									
Qu1	72,601 (21.72)	40,813 (22.48)	18,626 (22.46)	13,162 (18.88)		8,505 (24.97)	4,862 (24.94)	2,354 (25.66)	1,289 (23.92)
Qu2	70,937 (21.22)	39,614 (21.82)	17,778 (21.44)	13,545 (19.43)		8,092 (23.76)	4,636 (23.78)	2,217 (24.16)	1,239 (23.00)
Qu3	68,974 (20.64)	37,969 (20.91)	17,106 (20.63)	13,899 (19.93)		7,058 (20.72)	4,128 (21.17)	1,836 (20.01)	1,094 (20.30)
Qu4	64,640 (19.34)	34,748 (19.14)	16,005 (19.30)	13,887 (19.92)		6,125 (17.98)	3,446 (17.68)	1,679 (18.30)	1,000 (18.56)
Qu5	56,671 (16.96)	28,230 (15.55)	13,292 (16.03)	15,149 (21.73)		4,248 (12.47)	2,405 (12.34)	1,078 (11.75)	765 (14.20)
Health behaviours						**Health behaviours**			
Alcohol binge (frequency of consuming ≥6 units)									
n (%) missing	233,298 (69.80)	124,677 ()	55,235 (66.61)	53,386 (68.66)		12,053 (35.39)	6,847 (35.12)	3,183 (34.69)	2,023 (37.55)
n (%) in each category									
Never	50,329 (49.87)	27,890 (49.01)	13,688 (49.45)	8,751 (53.57)		10,214 (29.99)	5,800 (29.75)	2,741 (29.87)	1,673 (31.05)
Less than monthly	24,772 (24.54)	14,030 (24.65)	6,978 (25.21)	3,764 (23.04)		5,760 (16.91)	3,273 (16.79)	1,640 (17.87)	847 (15.72)
Monthly	8,998 (8.92)	5,287 (9.29)	2,492 (9.00)	1,219 (7.46)		2,122 (6.23)	1,275 (6.54)	566 (6.17)	281 (5.22)
Weekly	13,304 (13.18)	7,835 (13.77)	3,513 (12.69)	1,956 (11.97)		3,146 (9.24)	1,895 (9.72)	818 (8.92)	433 (8.04)
Daily or almost daily	3,526 (3.49)	1,868 (3.28)	1,011 (3.65)	647 (3.96)		763 (2.24)	405 (2.08)	227 (2.47)	131 (2.43)
Alcohol frequency n (%) missing	290 (0.09)	99 (0.05)	39 (0.05)	152 (0.22)		6 (0.02)	3 (0.02)	1 (0.01)	2 (0.04)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
n (%) in each category									
Daily/almost daily	71,942 (21.52)	40,332 (22.21)	18,523 (22.34)	13,087 (18.77)		8,092 (23.76)	4,600 (23.60)	2,289 (24.95)	1,203 (22.33)
3-4 times per week	80,933 (24.23)	46,916 (25.84)	20,272 (24.45)	13,805 (19.80		9,816 (28.82)	5,941 (30.47)	2,563 (27.93)	1,212 (24.35)
1-2 times per week	87,653 (26.23)	48,286 (26.59	21,469 (25.89)	17,898 (25.67)		8,660 (25.43)	5,000 (25.65)	2,314 (25.22)	1,346 (24.98)
1-3 times per month	36,751 (11.00)	18,870 (10.39)	9,405 (25.89)	8,476 (12.16)		3,520 (10.34)	1,885 (9.67)	967 (10.54)	668 (12.40
Special occasions	34,947 (10.46)	16, 716 (9.21)	8,401 (10.13)	9,830 (14.10)		2,552 (7.49)	1,300 (6.67)	686 (7.48)	566 (10.50)
Former	11,378 (3.40)	5,460 (3.01)	2,491 (3.00)	3,427 (4.92)		714 (2.10)	296 (2.03)	175 (1.91)	143 (2.65)
Never	10,273 (3.07)	3,048 (4.37)	2,317 (2.79)	3,048 (4.37)		698 (2.05)	370 (1.90)	180 (1.96)	148 (2.75)
Energy intake (KJ on previous day)									
n (%) missing	189,083 (56.57)	100,172 (55.16)	44,135 (53.23)	44,776 (64.22)		11,859 (34.58)	19,495 (33.77)	3,166 (34.51)	2,109 (39.14)
Mean (SD)	8898.03 (3022.75)	9010.12 (3090.70)	8825.39 (2893.99)	8645 (2975.42)		8988.99 (2941.18)	9092.83 (304.55)	8916.84 (2786.60)	8712.32 (2820.03)
Salt intake (added to food)									
n (%) missing	37 (0.01)	9 (0.00)	10 (0.01)	18 (0.03)		1 (0.01)	1 (0.01)	0 (0)	0 (0)
n (%) in each category									
Never/rarely	189,549 (56.71)	104,889 (57.76)	47,369 (57.13)	37,291 (53.48)		20,621 (60.55)	11,949 (61.29)	5,567 (60.68)	3,105 (57.63)
Sometimes	92,412 (27.65)	49,925 (27.49)	22,985 (27.72)	19,502 (27.97)		8,940 (26.25)	5,065 (25.98)	2,403 (26.19)	1,472 (27.32)
Usually	37,826 (11.32)	19,820 (10.91)	9,335 (11.26)	8,671 (12.44)		3,516 (10.32)	1,963 (10.07)	938 (10.22)	615 (11.41)
Always	14,403 (4.31)	6,944 (3.82)	3,218 (3.88)	4,241 (6.08)		980 (2.88)	518 (2.66)	266 (2.90)	196 (3.64)
Smoking status									
n (%) missing	1,138 (0.34)	470 (0.26)	208 (0.25)	460 (0.66)		65 (0.19)	31 (0.16)	13 (0.14)	21 (0.39)
Ever smoker, n (%) answered yes	38,344 (44.39)	81,163 (44.70)	36,504 (44.02)	33,292 (47.75)		13,376 (39.27)	7,787 (39.94)	3,457 (37,68)	2,132 (39.57)
Cardiovascular risk and biomarkers						**Cardiovascular risk and biomarkers**			

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Adiposity (BMI) n (%) missing Mean (SD)	 1,080 (0.32) 27.39 (4.75)	 391 (0.22) 26.85 (4.33)	 219 (0.26) 27.65 (4.86)	 470 (0.67) 28.49 (5.42)		 38 (0.11) 26.63 (4.26)	 26 (0.13) 26.22 (3.94)	 6 (0.07) 26.90 (4.42)	 6 (0.11) 27.62 (4.84)
Adiposity (waist circumference, cm) n (%) missing Mean (SD)	 563 (0.17) 90.32 (14.50)	 214 (0.12) 88.86 (12.78)	 118 (0.14) 91.33 (13.61)	 231 (0.33) 92.95 (14.51)		 13 (0.04) 88.30 (12.67)	 8 (0.04) 87.19 (12.17)	 4 (0.04) 90.38 (12.88)	 1 (0.02) 90.51 (13.60)
Arterial stiffness (stiffness index, higher = stiffer) n (%) missing Median (Q1, Q3)	 225,115 (67.35) 9.06 (6.91, 11.23)	 120,413 (66.31) 8.95 (6.82, 11.17)	 56,120 (67.68) 9.11 (7.00, 11.25)	 48,582 (69.68) 9.30 (7.10, 11.38)		 22,756 (66.82) 8.90 (6.78, 11.11)	 12,872 (66.03) 8.78 (6.72, 11.01)	 6,164 (67.18) 8.99 (6.80, 11.24)	 3,720 (69.04) 9.13 (6.99, 11.20)
Calcium (mmol/L) n (%) missing Mean (SD)	 42,538 (12.73) 2.38 (0.09)	 23,017 (12.68) 2.38 (0.09)	 10,513 (12.68) 2.38 (0.09)	 9,008 (12.92) 2.38 (0.10)		 4,496 (13.20) 2.38 (0.10)	 2,551 (13.09) 2.38 (0.09)	 1,141 (12.44) 2.37 (0.09)	 804 (14.92) 2.38 (0.09)
Total cholesterol (mmol/L) n (%) missing Mean (SD)	 15,577 (4.66) 5.71 (1.14)	 8,116 (4.47) 5.72 (1.12)	 3,919 (4.73) 5.71 (1.15)	 3,542 (5.08) 5.69 (1.19)		 1,666 (4.89) 5.72 (1.09)	 877 (4.50) 5.72 (1.08)	 5,388 (6.83) 5.72 (1.10)	 368 (6.83) 5.74 (1.09)
Diastolic BP (mmHg) n (%) missing Mean (SD)	 22,054 (6.60) 82.24 (10.66)	 11,684 (6.43) 82.00 (10.58)	 5.386 (6.50) 82.42 (10.70)	 4,984 (7.15) 82.75 (10.78)		 2,227 (6.54) 81.52 (10.42)	 1,231 (6.31) 81.20 (10.30)	 585 (6.38) 81.80 (10.45)	 411 (7.63) 82.17 (10.70)
Inflammation, CRP (mg/L) n (%) missing Median (Q1, Q3)	 16,266 (4.87) 1.32 (0.66, 2.73)	 8,449 (4.65) 1.18 (0.60, 2.40)	 4,120 (4.97) 1.38 (0.68, 2.87)	 3,697 (5.30) 1.70 (0.82, 3.55)		 1,744 (5.12) 1.08 (0.55, 2.17)	 912 (4.68) 0.99 (0.52, 1.96)	 447 (4.87) 1.16 (0.60, 2.35)	 385 (7.15) 1.32 (0.66, 2.69)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Systolic BP (mmHg) n (%) missing Mean (SD)	 22,059 (6.60) 140.15 (19.65)	 11,687 (6.44) 140.13 (19.64)	 5,387 (6.50) 139.37 (19.52)	 4,985 (7.15) 140.90 (19.80)		 2,227 (6.54) 137.52 (18.85)	 1,231 (6.31) 137 (18.85)	 585 (6.38) 136.88 (18.65)	 411 (7.63) 137.52 (18.85)
Medical diagnoses						**Medical diagnoses**			
Atrial fibrillation n (%) missing n (%) with diagnosis	 0 (0) 5,822 (1.74)	 0 (0) 2,919 (1.61)	 0 (0) 1,479 (1.78)	 0 (0) 1,424 (2.04)		 0 (0) 433 (1.27)	 0 (0) 246 (1.26)	 0 (0) 114 (1.24)	 0 (0) 73 (1.35)
Cardiovascular disease n (%) missing n (%) with diagnosis	 0 (0) 30,856 (9.15)	 0 (0) 15,173 (8.36)	 0 (0) 7,399 (8.92)	 0 (0) 8,014 (11.49)		 0 (0) 2,197 (6.45)	 0 (0) 1,218 (6.25)	 0 (0) 611 (6.66)	 0 (0) 367 (6.81)
Cerebrovascular disease n (%) missing n (%) with diagnosis	 0 (0) 5,280 (1.58)	 0 (0) 2,635 (1.30)	 0 (0) 1,194 (1.44)	 0 (0) 1,721 (2.47)		 0 (0) 307 (0.90)	 0 (0) 154 (0.79)	 0 (0) 92 (1.00)	 0 (0) 61 (1.13)
Chronic kidney disease n (%) missing n (%) with diagnosis	 0 (0) 3,882 (1.16)	 0 (0) 1,783 (0.98)	 0 (0) 952 (1.15)	 0 (0) 1,147 (1.16)		 0 (0) 261 (0.77)	 0 (0) 142 (0.73)	 0 (0) 66 (0.72)	 0 (0) 53 (0.98)
Chronic lung disease n (%) missing n (%) with diagnosis	 0 (0) 3,548 (1.06)	 0 (0) 1,328 (0.73)	 0 (0) 825 (0.99)	 0 (0) 1,395 (2.00)		 0 (0) 160 (0.47)	 0 (0) 75 (0.38)	 0 (0) 39 (0.43)	 0 (0) 46 (0.85)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Dementia n (%) missing n (%) with diagnosis	0 (0) 45 (0.01)	0 (0) 20 (0.01)	0 (0) 13 (0.02)	0 (0) 12 (0.02)		0 (0) 0 (0)	0 (0) 0 (0)	0 (0) 0 (0)	0 (0) 0 (0)
Diabetes n (%) missing n (%) with diagnosis	0 (0) 15,890 (4.75)	0 (0) 6,934 (3.82)	0 (0) 4,090 (4.93)	0 (0) 4,866 (6.98)		0 (0) 992 (2.91)	0 (0) 447 (2.29)	0 (0) 302 (3.29)	0 (0) 243 (4.51)
Head injury n (%) missing n (%) with diagnosis	0 (0) 1,486 (0.44)	0 (0) 824 (0.45)	0 (0) 325 (0.39)	0 (0) 337 (0.44)		0 (0) 95 (0.28)	0 (0) 54 (0.28)	0 (0) 27 (0.29)	0 (0) 14 (0.26)
Mood disorder n (%) missing n (%) with diagnosis	0 (0) 28,940 (8.66)	0 (0) 13,777 (7.59)	0 (0) 7,412 (8.94)	0 (0) 7,751 (11.12)		0 (0) 3,376 (9.91)	0 (0) 1,726 (8.85)	0 (0) 961 (10.47)	0 (0) 689 (12.79)
Musculoskeletal condition n (%) missing n (%) with diagnosis	0 (0) 144,216 (43.15)	0 (0) 76,561 (42.16)	0 (0) 34,175 (41.22)	0 (0) 33,480 (48.02)		0 (0) 14,132 (41.49)	0 (0) 8,136 (41.73)	0 (0) 3,713 (40.47)	0 (0) 2,283 (42.37)
Neurological condition n (%) missing n (%) with diagnosis	0 (0) 48,794 (14.60)	0 (0) 24,045 (13.24)	0 (0) 12,097 (14.59)	0 (0) 12,652 (18.15)		0 (0) 5,138 (15.09)	0 (0) 2,744 (14.08)	0 (0) 1,455 (15.86)	0 (0) 939 (17.43)
Psychotic conditions n (%) missing n (%) with diagnosis	0 (0) 779 (0.23)	0 (0) 299 (0.16)	0 (0) 208 (0.25)	0 (0) 272 (0.39)		0 (0) 31 (0.09)	0 (0) 21 (0.11)	0 (0) 7 (0.08)	0 (0) 3 (0.06)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]		Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Mental health						**Mental health**			
Current depression score n (%) missing Median (Q1, Q3)	31,179 (9.33) 1.54 (2.04)	13,661 (7.52) 1.34 (1.85)	6,569 (7.52) 1.62 (2.03)	10,949 (15.70) 1.97 (2.44)		2,100 (6.17) 1.00 (0.00, 2.00)	1,078 (5.53) 1.00 (0.00, 2.00)	493 (5.37) 1.00 (0.00, 2.00)	529 (9.82) 1.00 (0.00, 3.00)
Neuroticism score n (%) missing Median (Q1, Q3)	62,404 (18.67) 4.10 (3.25)	29,652 (16.33) 3.91 (3.20)	14,363 (17.32) 4.18 (3.23)	18,389 (26.37) 4.56 (3.40)		5,254 (15.43) 3.00 (1.00, 6.00)	2,776 (14.24) 3.00 (1.00, 6.00)	1,369 (14.92) 3.00 (1.00, 6.00)	1,109 (20.58) 3.00 (1.00, 7.00)
Traumatic events n (%) missing n (%) with ≥1 traumatic events	225,491 (67.47) 49,491 (45.51)	120,692 (66.47) 27,690 (45.47)	53,243 (64.21) 13,286 (44.77)	51,556 (74.94) 8,515 (46.87)		10,687 (31.38) 10,564 (31.02)	6,146 (31.53) 6,081 (31.19)	2,808 (30.60) 2,809 (30.62)	1,733 (32.16) 1.674 (31.07)
Worrier status n (%) missing n (%) answered yes	8,382 (2.51) 184,878 (55.32)	3,964 (2.18) 97,082 (53.46)	1,915 (2.31) 46,738 (56.37)	2,503 (3.59) 41,058 (58.89)		730 (2.14) 17,496 (51.37)	386 (1.98) 9,639 (49.44)	194 (2.11) 4,848 (52.84)	150 (2.78) 3,009 (55.85)
Medication						**Medication**			
Antihypertensive medication n (%) missing Any meds, n (%)	2,251 (0.67) 69,374 (20.76)	847 (0.47) 34,004 (18.73)	394 (0.48) 17, 318 (20.89)	1,010 (1.45) 18,052 (25.89)		272 (0.80) 4,210 (12.36)	76 (0.39) 2,258 (11.58)	33 (0.36) 1.170 (12.75)	163 (3.03) 782 (14.51)
Psychotropic medication n (%) missing Any meds, n (%)	9,813 (2.94) 28,834 (8.63)	5,120 (2.82) 12,687 (6.99)	2,408 (2.90) 7,158 (8.63)	2,285 (3.28) 8,989 (12.89)		942 (2.77) 1,957 (5.75)	519 (2.66) 932 (4.78)	262 (2.86) 528 (5.75)	161 (2.99) 497 (9.22)

Abbreviations: BMI, Body Mass Index; BP, blood pressure; CF, cognitive function; cm, centimetres; CRP, C-reactive protein; KJ, kilojoules; MET, Metabolic Equivalent of Task; mg/L, milligrams per litre; mmHg, millimeters of mercury; mmol/L, millimoles per litre; mvPA, moderate-vigorous PA; PA , physical activity; Q, quartile; Qu, Quintile; SD, Standard deviation

Footnotes for table:
a: Individuals were classified as active if they met ≥10 MET hours of moderate to vigorous PA per week. However, they also reported levels of light PA (walking) which did not contribute to this classification. Total PA includes light PA as well as mvPA. Therefore, there is a subset of individuals who are non-missing on light PA, but missing on both moderate and vigorous PA. These individuals will have a value for total PA but be counted as missing PA classification.
b: Global CF = mean of z scores on four tests (assuming at least two non-missing values)
c: Medical diagnoses based on linked health records with positive diagnosis indicating diagnosis on or before baseline assessment date

Table 1b: Cognitive outcomes for imaging subsample at follow-up

	Total sample	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**
n (%) of sample)	34,058 (100.00)	19,495 (57.24)	9,175 (26.94)	5,338 (15.82)
Original cognitive tests at follow-up				
Numeric memory (Z score) n (%) missing Mean (SD)	 10,573 (31.04) -0.38 (0.95)	 6,114 (31.36) -0.39 (0.95)	 2,718 (29.62) -0.42 (0.95)	 1,741 (32.31) -0.38 (0.95)
Pairs matching (Z score) n (%) missing Mean (SD)	 2,587 (7.60) 0.25 (1.06)	 1,527 (7.83) 0.24 (1.06)	 598 (6.52) 0.27 (1.05)	 463 (8.57) 0.24 (1.05)
Prospective memory n (%) missing n (%) correct on first attempt	 2,065 (6.06) 26,839 (78.80)	 1,225 (6.28) 15,810 (77.87)	 473 (5.16) 7,483 (81.56)	 367 (6.81) 4,176 (77.51)
Reaction time (Z score) n (%) missing Mean (SD)	 2,259 (6.63) 0.04 (0.96)	 1,339 (6.87) 0.05 (0.96)	 518 (5.65) 0.03 (0.95)	 402 (7.46) -0.01 (0.97)

	Total sample	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**
Reasoning (Z score) n (%) missing Mean (SD)	 2,638 (7.75) -0.20 (0.96)	 1,571 (8.06) - 0.22 (0.95)	 598 (6.52) - 0.10 (0.97)	 469 (8.70) -0.26 (0.98)
Additional follow-up tests				
Matrix pattern completion (Z score) n (%) missing Mean (SD)	 11,118 (32.64) -0.19 (0.94)	 6,417 (32.92) -0.21 (0.92)	 2,868 (31.26) -0.12 (0.95)	 1,833 (34.02) -0.25 (0.94)
Paired associate learning (Z score) n (%) missing Mean (SD)	 11,581 (34.00) -0.22 (0.83)	 6,702 (34.38) -0.23 (0.83)	 2,994 (32.63) - 0.17 (0.81)	 1,885 (34.99) -0.29 (0.85)
Symbol Digit Substitution (Z score) n (%) missing Mean (SD)	 11,097 (32.58) -0.08 (0.96)	 6,410 (32.88) -0.10 (0.94)	 2,858 (31.15) -0.03 (0.94)	 1,829 (33.95) -0.12 (0.96)
Tower rearranging (Z score) n (%) missing Mean (SD)	 11,311 (33.21) -0.12 (0.94)	 6,516 (33.42) -0.13 (0.93)	 2,927 (31.90) -0.09 (0.96)	 1,868 (34.67) -0.14 (0.96)
Trails B-A, time (Z score) n (%) missing Mean (SD)	 11,730 (34.44) 0.03 (0.96)	 6,795 (34.86) 0.00 (0.96)	 2,996 (32.65) 0.08 (0.94)	 1,939 (35.99) 0.03 (0.95)

	Total sample	**Active (≥10 MET hours mvPA/week)**	**Inactive (<10 MET hours of mvPA/week)**	**Missing PA status[a]**
Trails A, time (Z score) n (%) missing Mean (SD)	 11,134 (32.69) 0.04 (0.96)	 6,435 (33.01) 0.02 (0.96)	 2,863 (31.20) 0.08 (0.95)	 1,836 (34.08) 0.05 (0.95)
Trails B, time (Z score) n (%) missing Mean (SD)	 11,730 (34.44) 0.04 (0.96)	 6,795 (34.86) 0.00 (0.96)	 2,996 (32.65) 0.10 (0.95)	 1,939 (35.99) 0.04 (0.95)
Trails B, errors (Z score) n (%) missing Mean (SD)	 11,230 (32.97) 0.99 (1.57)	 6,493 (33.31) 0.96 (1.57)	 2,885 (31.44) 1.05 (1.55)	 1,852 (34.37) 0.96 (1.57)
Composite CF measures				
Global CF (Z score)[b] n (%) missing Mean (SD)	 2,370 (6.96) -0.05 (0.56)	 1,407 (7.22) -0.06 (0.57)	 539 (5.87) -0.10 (0.56)	 424 (7.87) -0.05 (0.57)
Processing speed composite (Z score)[c] n (%) missing Mean (SD)	 11,168 (32.79) -0.03(0.72)	 6,446 (33.06) -0.06 (0.74)	 2,879 (31.38) -0.01 (0.73)	 1,843 (34.21) -0.06 (0.74)
Executive function composite (Z score)[d] n (%) missing Mean (SD)	 11,186 (32.84) -0.02 (0.72)	 6,461 (33.14) -0.04 (0.72)	 2,875 (31.34) 0.03 (0.73)	 1,850 (34.34) -0.02 (0.72)

	Total sample	Active (≥10 MET hours mvPA/week)	Inactive (<10 MET hours of mvPA/week)	Missing PA status[a]
Reasoning composite (Z score)[e] n (%) missing Mean (SD)	 11,281 (33.12) -0.21 (0.79)	 6,158 (33.43) - 0.23 (0.77)	 2,905 (31.66) -0.12 (0.79)	 1,858 (34.48) -0.26 (0.81)
Memory composite (Z score)[f] n (%) missing Mean (SD)	 10,410 (30.57) -0.12 (0.62)	 6,024 (30.90) -0.13 (0.62)	 2,681 (29.22) -0.08 (0.61)	 1,705 (31.64) -0.12 (0.62)

Abbreviations: CF, cognitive function; MET, Metabolic Equivalent of Task; mvPA, moderate-vigorous PA; PA , physical activity; SD, Standard deviation

Footnotes for table:

a: Individuals were classified as active if they met ≥10 MET hours of moderate to vigorous PA per week. However, they also reported levels of light PA (walking) which did not contribute to this classification. Total PA includes light PA as well as mvPA. Therefore, there is a subset of individuals who are non-missing on light PA, but missing on both moderate and vigorous PA. These individuals will have a value for total PA but be counted as missing PA classification

b: Global CF = mean of z scores on ten tests assuming at least two non-missing values

c: Processing speed composite = mean of Digit Symbol Substitution and Reaction Time (assuming non-missing on both measures)

d: Executive function composite = mean of Tower Rearranging, Trails A and Trails B completion time (assuming non-missing on two measures)

e: Reasoning composite = mean of Reasoning test and Matrix Pattern Completion (assuming non-missing on both measures)

f: Memory composite = mean of Pairs Matching, Numeric Memory and Paired Associate Learning (assuming non -missing on two measures)

Table 2: Cross-sectional regression models for baseline cognitive function

Exposure	Cognitive score	Unadjusted					Adjusted[b]				
		n	Estimate[a]	95% CI	*p* (uncorr)	*p* (FDR)	*n*	Estimate	95% CI	*p* (uncorr)	*p* (FDR)[c]
Total PA, self-report (MET hrs/week)	Reaction Time	305,294	.000131	.0000659, .0001962	.0001	.0001	29,810	-.000168	-.0004764, .0001403	.4471	.5365
	Pairs Matching	300,847	-.0005355	-.0006077, -.004634	<.0001	<.0001	29,664	-.0001841	-.0004921, .0001239	.2855	.5365
	Reasoning	100,204	-.0023377	-.0024488, -.0022266	<.0001	<.0001	12,438	-.0009303	-.001292, -.0005685	<.0001	<.0001
	Numeric Memory	31,854	-.0012581	-.014492, -.0010669	<.0001	<.0001	3,613	-.0001427	-.0008524, .0005669	.6934	.6934
	Global CF[d]	300,915	-.0004846	-.0005331,-.0004362	<.0001	<.0001	29,695	-.0001974	-.000393, -.00000017	.0480	.1440
	Prospective Memory	89,022	.9981991	.9979159, .998482	<.0001	<.0001	2,548	.9986968	.9955481, 1.001856	.4183	.53652

Abbreviations: CF, cognitive function; CI, confidence interval; FDR, False Discovery Rate; MET, Metabolic Equivalent of Task; PA , physical activity; Uncorr, uncorrected

a: All expressed as z scores, except Prospective Memory which is expressed as an odds ratio

b: Adjusted for: alcohol binge, alcohol frequency, antihypertensive medication, apoe-e4 allele count, BMI, cardiovascular disease dx, dementia genetic risk score, diabetes dx, distance to major road, friend and family visits, gender, hdl cholesterol, head injury dx, household income, kidney disease dx, KJ of energy, ldl cholesterol, living alone status, manual work, mood disorder dx, musculoskeletal dx, neurological disorder dx, neuroticism score psychosis dx, psychotropic medication, salt added to food, smoking status, Townsend deprivation score, trauma status, waist circumference, worrier status. Also adjusted for technical covariates used with genetic risk scores

c: Probability adjusted using the Simes-Benjamini-Hochberg method implemented in the Stata qqvalue package

d: Global CF = mean of z scores on four tests (assuming at least two non-missing values)

Table 3: Longitudinal regression models for follow-up cognitive function

Exposure	Cognitive score	Unadjusted					Adjusted[b]				
		n	Estimate[a]	95% CI	p (uncorr)	p (FDR)[b]	n	Estimate	95% CI	p (uncorr)	p (FDR)[c]
Total PA, self-report (MET hrs/week)	Reaction Time	30,153	-.0000553	-.0002922, .0001816	.6474	.6474	6,816	.0001664	-.0004255, .0007584	.5816	.6394
	Pairs Matching	29,845	-.000428	-.0006911, -.0001649	.0014	.0015	6,780	-.0001783	-.000858, .0005014	.6071	.6394
	Reasoning	29,801	-.0027912	-.0030282, -.0025542	<.0001	<.0001	6,779	-.0017126	-.0022898,-.001134	<.0001	<.0001
	Numeric Memory	22,321	-.001635	-.0019111, -.0013588	<.0001	<.0001	5,006	-.0015375	-.0022407,-.0008344	.6394	.6394
	Symbol Digit Substitution	21,831	-.0016325	-.0019087, -.0013564	<.0001	<.0001	4,886	-.0006917	-.0014112, .0000277	.0595	.0893
	Paired Associate Learning	21,343	-.0015357	-.0017798, -.0012917	<.0001	<.0001	4,805	-.0004867	-.0011079, .0001346	.1247	.1727
	Tower Rearranging	21,626	-.0011721	-.0014512, -.0008931	<.0001	<.0001	4,865	-.0004471	-.0011614, .0002672	.2198	.2826
	Matrix Pattern Completion	21,804	-.0019887	-.0022625, -.001715	<.0001	<.0001	4,881	-.0009316	-.0016148, -.0002483	.0075	.0193
	Prospective Memory	30,330	.9971918	.9965854 .9977987	<.0001	<.0001	6,840	.9981081	.9961615, 1.000058	.0573	.0893
	Trails A (time)	21,709	-.0013174	-.0015988,- .0010359	<.0001	<.0001	4,874	-.0007286	-.0014642,.0000005	.0518	.0893

	Trails B (time)	21,225	-.0019304	- .0022155,-.0016453	<.0001	<.0001	4,827	-.0010314	-.0017477, -.0003152	.0048	.0173
	Trails B-A (time)	21,225	-.0015662	-.0018518,-.0012806	<.0001	<.0001	4,827	-.0009099	-.0016336, -.0001862	.0137	.0274
	Trails B (errors)	21,698	-.002139	-.0026008, -.0016773	<.0001	<.0001	4,866	-.0015567	.-0027763, -.000337	.0124	.0274
	Processing Speed (comp)[d]	21,767	-.0009248	-.0011378, -.0007117	<.0001	<.0001	4,873	-.0003291	-.0008781, .0002199	.2400	.2880
	Executive Function (comp)[e]	21,742	-.001489	-.0017003 , -.0012777	<.0001	<.0001	4,871	-.0007267	-.0012586 -.0001947	.0074	.0193
	Reasoning (comp)[f]	21,655	-.0024415	-.0026714, -.0022116	<.0001	<.0001	4,866	-.0013687	-.0019174, -.0008199	<.0001	<.0001
	Memory (comp)[g]	22,467	-.0011893	-.0013674, -.0010112	<.0001	<.0001	5,029	-.0007394	-.0011853, -.0002936	.0012	.0054
	Global CF (comp)[h]	30,048	-.0012926	-.0014324, -.0011529	<.0001	<.0001	6,804	-.0006555	-.0010004, -.0003107	.0002	.0012
Exposure	**Cognitive score**	**Unadjusted**					**Adjusted**				
Physical Activity, accelerometery		**n**	**Estimate[a]**	**95% CI**	**p (uncorr)**	**p (FDR)**	**n**	**Estimate**	**95% CI**	**p (uncorr)**	**p (FDR)**
	Reaction Time	14,307	.0029557	.0012409, .0046705	.0007	.0063	3,935	.002567	-.0007845, .0059099	.1334	.8058
	Pairs Matching	14,164	-.0009559	-.0028609, .0009492	.3254	.3584	3,919	.0006546	-.0046161, .0033069	.7460	.8952
	Reasoning	14,148	.0009323	-.0026586, .000794	.2898	.3478	3,919	.0001802	-.0035365, .0031761	.9162	.9162
	Numeric Memory	9,901	.0019117	-.0001713, .0039947	.0720	.1566	2,837	.000409	-.003878, .0046959	.8516	.9162

(milligravity units)	Symbol Digit Substitution	9,683	.0030884	.0010321, .0051447	.0032	.0162	2,772	-.0018307	-.0059394, .0022781	.3824	.8952
	Paired Associate Learning	9,522	.0015174	-.0003052, .0033401	.1027	.1849	2,731	-.0006999	-.0042253, .0028255	.6971	.8952
	Prospective Memory	14,392	1.006038	1.0007, 1.011405	.0266	.0798	3,828	1.003531	.9898692, 1.017381	.6143	.8952
	Tower Rearranging	9,611	-.0019001	-.0039706, .0001704	.0721	.1566	2,759	-.0028300	-.0070429, .0013830	.1879	.8058
	Matrix Pattern Completion	9,670	-.0009633	-.0030134, .0010868	.3570	.3584	2,768	-.0009649	-.0051264, .0031966	.6494	.8952
	Trails A (time)	9,668	.0030946	.0010132, .005176	.0036	.0162	2,766	-.0010388	-.0052679, .0031902	.6301	.8952
	Trails B (time)	9,455	.0026019	.0004972, .0047067	.0154	.0554	2,742	-.0026095	-.0067544, .0015354	.2171	.8058
	Trails B-A (time)	9,455	.0018798	-.0002126, .0039721,	.0783	.1566	2,742	-.0023665	-.006514, .0017809	.2633	.8058
	Trails B (errors)	9,362	.001587	-.001800, .004974	.3584	.3584	2,760	-.0051376	-.0117453, .0014702	.1275	.8058
	Processing Speed (comp)[d]	9,653	.0035774	.0019994, .0051555	<.0001	.0002	2,763	.0005342	-.0026049, .0036733	.7386	.8952
	Executive Function (comp)[e]	9,650	.0001975	-.0029336, .0033287	.1309	.2142	2,764	.0001975	-.0029336, .0033287	.9016	.9162
	Reasoning (comp)[f]	9,612	-.0010418	-.0027634, .0006798	.2356	.3228	2,763	.0016841	-.0051657, .0014390	.2686	.8058
	Memory (comp)[g]	9,964	.0009278	-.0004027, .0022584	.1717	.2576	2,848	-.0006474	-.0032057, .0019109	.6198	.8952
	Global CF (comp)[h]	14,284	.0005975	-.0004228, .0016178	.2510	.3227	3,933	.0007286	-.0012778, .0027349	.4765	.8952

Abbreviations: CF, cognitive function; CI, confidence interval; FDR, False Discovery Rate; MET, Metabolic Equivalent of Task; PA , physical activity; Uncorr, uncorrected
a: All expressed as z scores, except Prospective Memory which is expressed as an odds ratio
b: Adjusted for: alcohol binge, alcohol frequency, antihypertensive medication, apoe-e4 allele count, BMI, cardiovascular disease dx, dementia genetic risk score, diabetes dx, distance to major road, friend and family visits, gender, hdl cholesterol, head injury dx, household income, kidney disease dx, KJ of energy, ldl cholesterol, living alone status, manual work, mood disorder dx, musculoskeletal dx, neurological disorder dx, neuroticism score psychosis dx, psychotropic medication, salt added to food, smoking status, Townsend deprivation score, trauma status, waist circumference, worrier status. Also adjusted for technical covariates used with genetic risk scores.
c: Probability adjusted using the Simes-Benjamini-Hochberg method implemented in the Stata qqvalue package.
d: Processing speed composite = mean of Digit Symbol Substitution and Reaction Time (assuming non-missing on both measures)
e: Executive function composite = mean of Tower Rearranging, Trails A and Trails B completion time (assuming non-missing on two measures)
f: Reasoning composite = mean of Reasoning test and Matrix Pattern Completion (assuming non-missing on both measures)
g: Memory composite = mean of Pairs Matching, Numeric Memory and Paired Associate Learning (assuming non -missing on two measures)
h: Global CF = mean of z scores on ten tests (assuming at least two non-missing values)

Discussion

Main findings

In this study using a large cohort of middle to early old age adults of white British ancestry to estimate the causal effect of PA on CF, there was virtually zero association between these variables. Due to very large sample sizes some of the effect estimates did reach thresholds for statistical significance; however, they were of trivially small magnitude, and became smaller after adjustment for covariates. Planned analyses of mediating pathways via structural brain differences were not conducted due to the inconsistency in the direction of the total effects estimates. This pattern of results was unexpected as it does not align with most of the recent literature reviewed in chapter one, although a minority of the reviewed studies also reported no association between PA and CF (Morgan et al., 2012; Sabia et al., 2017; Tan et al., 2017; Verghese, Wang, Katz, Sanders, & Lipton, 2009). In common with two of these studies, the UK Biobank sample was younger at baseline than most of the other cohorts in which protective effects have been found. Taken together, the present findings may support the suggestion that observed effects in older baseline samples reflect reverse causality, whereby some participants had preclinical cognitive decline at baseline, and lower PA was a symptom, rather than a cause, of changes in brain health. Indeed, a growing body of evidence demonstrates that earlier changes in behaviour predict later dementia diagnosis, before cognitive symptoms manifest (Bayat et al., 2021). Studies with younger samples at baseline reduce the risk that preclinical disease processes have begun. Other potential explanations for the present finding are considered below.

Selection bias

The UK Biobank sample is known not to be representative of the general population, with participants less likely to engage in unhealthy behaviours and experience negative health outcomes (Fry et al., 2017). When both exposure and outcome are related to participation in studies this can lead to collider bias (discussed in figure 1, chapter one), and it is plausible that both higher levels of PA and better cognitive health influence participation and retention within UK Biobank. Indeed, the sample who returned for imaging were more active, less deprived and healthier than the total sample. The implication is that the true relationship between PA and CF may be distorted within UK Biobank due to participants' better health, as has been observed with other anticipated effects (Lyall et al., 2021).

Model misspecification

Another explanation may be model misspecification. In particular, the model in this thesis

represents PA (at the point it is measured in UK Biobank) as downstream of health outcomes such as cardiometabolic disease. As such, these were specified as confounders and adjusted for. However, one hypothesised mechanism by which PA may affect CF is by reducing cardiometabolic risk, meaning the adjusted models in this the present study would have removing part of the effect of interest by over-adjustment (Schisterman, Cole, & Platt, 2009). However, given that the unadjusted effect sizes were also trivially small it is unlikely that over-adjustment accounts for the observed findings. There were also three confounders specified by the model that were not adjusted for due to lack of matches within the available data (childhood PA, earlier adulthood PA and childhood IQ). Whilst their omission represents residual bias within the model, it remains unlikely that including these would have made a substantial difference to the pattern of results.

Age of sample

It may also be the case that the benefits of PA on CF would not be observable until later in life when a greater degree of cognitive decline would be expected (Salthouse, 2009). Indeed, in much of the literature reviewed in chapter one which found a protective effect of PA on CF, the samples were into the eighth and ninth decades of life at follow-up, whereas the present sample at follow-up were in their seventh decade of life. It remains possible that the protective effect of PA is yet to be realised within this relatively young cohort.

Strengths and limitations

There are several features of this study that represent both a strength and a weakness. The application of a causal inference framework to examine the effect of PA on CF represents a novel contribution to the literature, and the construction of the DAG was done to a standard of rigour and transparency that is not common within existing literature (Tennant et al., 2019). However, the complexities of the model posed practical limitations such as being unable to interrogate it comprehensively using the available software, meaning that the plausible variations of the specified model were not explored. Nor were the implied independencies of the model tested against the measured data, which is another way of assessing model fit (Textor, Hardt, & Knüppel, 2011).

The use of UK Biobank data also represents a trade-off in terms of strengths and limitations. The range and detail of measures available within this resource allowed a close approximation of the complex conceptual model to be estimated statistically, including genomic, environmental and lifestyle covariates. However, the internal and external validity of the

findings is limited due to the selection bias within the sample, and the genetic ancestry stratification means results cannot be generalised beyond populations of white British ancestry. A sensitivity analysis was performed to examine the effect of missing data, however further analyses involving multiple imputation of missing values were not conducted.

Finally, the measurement of exposure and outcome were strengths of the present study. Cognitive data was utilised to represent the various conceptualisations of CF within the literature. Previous studies have tended to examine either individual test scores, domain-level averages, or total-scores across domains (global CF), whereas the present study considered all these variations. This study also benefited from having both self-report and objective measures of PA.

Implications and future directions

Due to limitations to both internal and external validity discussed above, the present results should be interpreted with caution. However, in the context of existing literature, the finding of no meaningful association between PA and CF aligns with other studies that had younger baseline samples and may lend weight to the reverse causation hypothesis. Alternatively, the null findings may reflect the suppression of true effects due to collider bias induced by the factors influencing participation and retention within the UK Biobank sample. Future research using UK Biobank can explore whether the hypothesised protective effect of PA on CF does emerge as the cohort matures.

Conclusions

This study estimated the causal effect of PA on CF based on a comprehensive model derived within a causal inference framework, using a very large sample of middle to early old-aged UK adults of white British ancestry. The expected protective effect was not observed. This may reflects selection bias within UK Biobank, or the relatively young age of the sample at follow-up. However, it is also possible that previous studies with older samples have over-estimated the association of PA and CF due to reverse causation. Future research could utilise the conceptual model advanced in this thesis to examine whether effects of PA on CF emerge as the UK Biobank sample ages.

References

Batouli, S. A. H., & Saba, V. (2017). At least eighty percent of brain grey matter is modifiable by physical activity: A review study. *Behavioural brain research, 332*, 204-217.

Bayat, S., Babulal, G. M., Schindler, S. E., Fagan, A. M., Morris, J. C., Mihailidis, A., & Roe, C. M. (2021). GPS driving: a digital biomarker for preclinical Alzheimer disease. *Alzheimer's research & therapy, 13*(1), 1-9.

Blondell, S. J., Hammersley-Mather, R., & Veerman, J. L. (2014). Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. *BMC public health, 14*(1), 1-12.

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., . . . Chou, R. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *British journal of sports medicine, 54*(24), 1451-1462.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*: Cambridge University Press.

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

Celis-Morales, C. A., Lyall, D. M., Petermann, F., Anderson, J., Ward, J., Iliodromiti, S., . . . Pell, J. (2019). Do physical activity, commuting mode, cardiorespiratory fitness and sedentary behaviours modify the genetic predisposition to higher BMI? Findings from a UK Biobank study. *International Journal of Obesity, 43*(8), 1526-1538.

Cox, S., Ritchie, S., Fawns-Ritchie, C., Tucker-Drob, E., & Deary, I. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence, 76*, 101376.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., ... & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. Medicine & science in sports & exercise, 35(8), 1381-1395.

Cunningham, C., O'Sullivan, R., Caserotti, P., & Tully, M. A. (2020). Consequences of physical inactivity in older adults: A systematic review of reviews and meta‐analyses. *Scandinavian journal of medicine & science in sports, 30*(5), 816-827.

Deary, I. J. (2001). Human intelligence differences: a recent history. *Trends in cognitive sciences, 5*(3), 127-130.

Deary, I. J. (2020). *Intelligence: A very short introduction*: Oxford University Press.

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., . . . Owen, C. G. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. *PloS one, 12*(2), e0169649.

Erickson, K. I., Hillman, C. H., & Kramer, A. F. (2015). Physical activity, brain, and cognition. *Current opinion in behavioral sciences, 4*, 27-32.

Fawns-Ritchie, C., & Deary, I. J. (2020). Reliability and validity of the UK Biobank cognitive tests. *PloS one, 15*(4), e0231627.

Firth, J., Stubbs, B., Vancampfort, D., Schuch, F., Lagopoulos, J., Rosenbaum, S., & Ward, P. B. (2018). Effect of aerobic exercise on hippocampal volume in humans: a systematic review and meta-analysis. *Neuroimage, 166*, 230-238.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research, 12*(3), 189-198.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., . . . Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American journal of epidemiology, 186*(9), 1026-1034.

Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence, 64*, 18-29.

Guure, C. B., Ibrahim, N. A., Adam, M. B., & Said, S. M. (2017). Impact of physical activity on cognitive decline, dementia, and its subtypes: meta-analysis of prospective studies. *BioMed research international, 2017*.

Halloway, S., Wilbur, J., Schoeny, M. E., & Barnes, L. L. (2017). The relation between physical activity and cognitive change in older Latinos. *Biological research for nursing, 19*(5), 538-548.

Hamer, M., & Chida, Y. (2009). Physical activity and risk of neurodegenerative disease: a systematic review of prospective evidence. *Psychological medicine, 39*(1), 3.

Lambert, J. D., Greaves, C. J., Farrand, P., Haase, A. M., & Taylor, A. H. (2017). Development of a web-based intervention (eMotion) based on behavioural activation to promote physical activity in people with depression. *Mental Health and Physical Activity, 13*, 120-136.

Lear, S. A., Hu, W., Rangarajan, S., Gasevic, D., Leong, D., Iqbal, R., . . . Kumar, R. (2017). The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. *The Lancet, 390*(10113), 2643-2654.

Lyall, D., Quinn, T., Lyall, L., Ward, J., Anderson, J., Smith, D., … Cullen, B. (2021, July 27). Quantifying bias in psychological and physical health in the UK Biobank imaging sub-sample. https://doi.org/10.31234/osf.io/upvb9

Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., . . . Deary, I. J. (2016). Cognitive test scores in UK Biobank: data reduction in 480,416 participants and longitudinal stability in 20,346 participants. *PloS one, 11*(4), e0154222.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., ... & Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience*, *19*(11), 1523-1536.

Milton, D. C., Ward, J., Ward, E., Lyall, D. M., Strawbridge, R. J., Smith, D. J., & Cullen, B. (2021). The association between C-reactive protein, mood disorder, and cognitive function in UK Biobank. *European Psychiatry, 64*(1).

Morgan, G. S., Gallacher, J., Bayer, A., Fish, M., Ebrahim, S., & Ben-Shlomo, Y. (2012). Physical activity in middle-age and dementia in later life: findings from a prospective cohort of men in Caerphilly, South Wales and a meta-analysis. *Journal of Alzheimer's Disease, 31*(3), 569-580.

Norton, K., Norton, L., & Sadgrove, D. (2010). Position statement on physical activity and exercise intensity terminology. *Journal of science and medicine in sport, 13*(5), 496-502.

Penke, L., Maniega, S. M., Bastin, M., Hernández, M. V., Murray, C., Royle, N., . . . Deary, I. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular psychiatry, 17*(10), 1026-1030.

Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience & Biobehavioral Reviews, 57*, 411-432.

Puzo, C., Labriola, C., Sugarman, M. A., Tripodis, Y., Martin, B., Palmisano, J. N., . . . McKee, A. C. (2019). Independent effects of white matter hyperintensities on cognitive, neuropsychiatric, and functional decline: a longitudinal investigation using the National Alzheimer's Coordinating Center Uniform Data Set. *Alzheimer's research & therapy, 11*(1), 1-13.

Rajan, K. B., Barnes, L. L., Skarupski, K. A., De Leon, C. F. M., Wilson, R. S., & Evans, D. A. (2015). Physical and cognitive activities as deterrents of cognitive decline in a biracial population sample. *The American Journal of Geriatric Psychiatry, 23*(12), 1225-1233.

Ritchie, S. J., Bastin, M. E., Tucker-Drob, E. M., Maniega, S. M., Engelhardt, L. E., Cox, S. R., . . . Pattie, A. (2015). Coupled changes in brain white matter microstructure and fluid intelligence in later life. *Journal of Neuroscience, 35*(22), 8672-8682.

Ritchie, S. J., Booth, T., Hernández, M. d. C. V., Corley, J., Maniega, S. M., Gow, A. J., . . . Starr, J. M. (2015). Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence, 51*, 47-56.

Sabia, S., Dugravot, A., Dartigues, J.-F., Abell, J., Elbaz, A., Kivimäki, M., & Singh-Manoux, A. (2017). Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *Bmj, 357*.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of aging, 30*(4), 507-514.

Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology (Cambridge, Mass.), 20*(4), 488.

Sofi, F., Valecchi, D., Bacci, D., Abbate, R., Gensini, G. F., Casini, A., & Macchi, C. (2011). Physical activity and risk of cognitive decline: a meta‐analysis of prospective studies. *Journal of internal medicine, 269*(1), 107-117.

Ströhle, A. (2009). Physical activity, exercise, depression and anxiety disorders. *Journal of neural transmission, 116*(6), 777-784.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *Plos med, 12*(3), e1001779.

Tan, Z. S., Spartano, N. L., Beiser, A. S., DeCarli, C., Auerbach, S. H., Vasan, R. S., & Seshadri, S. (2017). Physical activity, brain volume, and dementia risk: the Framingham study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 72*(6), 789-795.

Tennant, P. W., Harrison, W. J., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., . . . Textor, J. (2019). Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. *medRxiv*.

Textor, J., Hardt, J., & Knüppel, S. (2011). DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology, 22*(5), 745.

Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., . . . Hayes, G. (2011). Quality control procedures for genome‐wide association studies. *Current protocols in human genetics, 68*(1), 1.19. 11-11.19. 18.

Vakil, E. (2012). Neuropsychological assessment: Principles, rationale, and challenges. *Journal of clinical and experimental neuropsychology, 34*(2), 135-150.

Verghese, J., Wang, C., Katz, M. J., Sanders, A., & Lipton, R. B. (2009). Leisure activities and risk of vascular cognitive impairment in older adults. *Journal of geriatric psychiatry and neurology, 22*(2), 110-118.

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & Initiative, S. (2014). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *International journal of surgery, 12*(12), 1495-1499.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson, 22*(498), 1.

World Health Organization. (2006). Global strategy on diet, physical activity and health: a framework to monitor and evaluate implementation.

APPENDICES

Systematic review

Appendix 1.1: Journal of Neuropsychology submission guidelines*

* *Including scope and manuscript requirements. For full guidance see link below:*

https://bpspsychub.onlinelibrary.wiley.com/hub/journal/17486653/homepage/forauthors.html)

JNP AUTHOR GUIDELINES

Sections

1. SUBMISSION

Authors should kindly note that submission implies that the content has not been published or submitted for publication elsewhere except as a brief abstract in the proceedings of a scientific meeting or symposium.

Once the submission materials have been prepared in accordance with the Author Guidelines, manuscripts should be submitted online at http://www.editorialmanager.com/jnp

Click here for more details on how to use **Editorial Manager**.

All papers published in the *Journal of Neuropsychology* are eligible for Panel A: Psychology, Psychiatry and Neuroscience in the Research Excellence Framework (REF).

Data protection:

By submitting a manuscript to or reviewing for this publication, your name, email address, and affiliation, and other contact details the publication might require, will be used for the regular operations of the publication, including, when necessary, sharing with the publisher (Wiley) and partners for production and publication. The publication and the publisher recognize the importance of protecting the personal information collected from users in the operation of these services, and have practices in place to ensure that steps are taken to maintain the security, integrity, and privacy of the personal data collected and processed. You can learn more at **https://authorservices.wiley.com/statements/data-protection-policy.html**.

Preprint policy:

This journal will consider for review articles previously available as preprints. Authors may also post the submitted version of a manuscript to a preprint server at any time. Authors are requested to update any pre-publication versions with a link to the final published article.

2. AIMS AND SCOPE

The Journal of Neuropsychology publishes original contributions to scientific knowledge in neuropsychology including:

- clinical and research studies with neurological, psychiatric and psychological patient populations in all age groups
- behavioural or pharmacological treatment regimes
- cognitive experimentation and neuroimaging
- multidisciplinary approach embracing areas such as developmental psychology, neurology, psychiatry, physiology, endocrinology, pharmacology and imaging science

The following types of paper are invited:

- papers reporting original empirical investigations
- theoretical papers; provided that these are sufficiently related to empirical data
- review articles, which need not be exhaustive, but which should give an interpretation of the state of research in a given field and, where appropriate, identify its clinical implications
- brief reports and comments
- case reports
- fast-track papers (included in the issue following acceptation) reaction and rebuttals (short reactions to publications in JNP followed by an invited rebuttal of the original authors)
- special issues.

3. MANUSCRIPT CATEGORIES AND REQUIREMENTS

- Research papers should be no more than 6000 words (excluding the abstract, reference list, tables and figures). Multiple citations for a single point are usually duplicative and authors are urged to cite the best reference. In exceptional cases the Editor retains discretion to publish papers beyond this length where the clear and concise expression of the scientific content requires greater length (e.g., explanation of a new theory or a substantially new method). Authors must contact the Editor prior to submission in such a case.
- Brief communications are short reports of original research or case reports. They are limited to a maximum of 1500 words (excluding the abstract, reference list, tables and figures) and have a total of up to three tables or figures, and no more than 10 references.
- Theoretical or review articles are full-length reviews of, or opinion statements regarding, the literature in a specific scientific area. They should be no more than 4000 words (excluding the abstract, reference list, tables and figures) and have no more than 45 references. Multiple citations for a single point are usually duplicative and authors are urged to cite the best reference. In exceptional cases the Editor retains discretion to publish papers beyond this length where the clear and concise expression of the scientific content requires greater length (e.g., explanation of a new theory or a substantially new method). Authors must contact the Editor prior to submission in such a case.
- Please refer to the separate guidelines for Registered Reports.
- All systematic reviews must be pre-registered.

4. PREPARING THE SUBMISSION

Free Format Submission

Journal of Neuropsychology now offers free format submission for a simplified and streamlined submission process.

Before you submit, you will need:

- Your manuscript: this can be a single file including text, figures, and tables, or separate files – whichever you prefer. All required sections should be contained in your manuscript, including abstract, introduction, methods, results, and conclusions. Figures and tables should have legends. References may be submitted in any style or format, as long as it is consistent throughout the manuscript. If the manuscript, figures or tables are difficult for you to read, they will also be difficult for the editors and reviewers. If your manuscript is difficult to read, the editorial office may send it back to you for revision.
- The title page of the manuscript, including a data availability statement and your co-author details with affiliations. *(Why is this important? We need to keep all co-authors informed of the outcome of the peer review process.)* You may like to use this template for your title page.

Important: the journal operates a double-blind peer review policy. Please anonymise your manuscript and prepare a separate title page containing author details. *(Why is this important? We need to uphold rigorous ethical standards for the research we consider for publication.)*

- An ORCID ID, freely available at https://orcid.org. *(Why is this important? Your article, if accepted and published, will be attached to your ORCID profile. Institutions and funders are increasingly requiring authors to have ORCID IDs.)*

To submit, login at **https://www.editorialmanager.com/jnp/default.aspx** and create a new submission. Follow the submission steps as required and submit the manuscript.

If you are invited to revise your manuscript after peer review, the journal will also request the revised manuscript to be formatted according to journal requirements as described below.

Revised Manuscript Submission

Contributions must be typed in double spacing. All sheets must be numbered.

Cover letters are not mandatory; however, they may be supplied at the author's discretion. They should be pasted into the 'Comments' box in Editorial Manager.

Parts of the Manuscript

The manuscript should be submitted in separate files: title page; main text file; figures/tables; supporting information.

Title Page

You may like to use **this template** for your title page. The title page should contain**:**

- A short informative title containing the major key words. The title should not contain abbreviations (see Wiley's best practice SEO tips);
- A short running title of less than 40 characters;
- The full names of the authors;
- The author's institutional affiliations where the work was conducted, with a footnote for the author's present address if different from where the work was conducted;
- Abstract;
- Keywords;
- Data availability statement (see Data Sharing and Data Accessibility Policy);
- Acknowledgments.

Authorship

Please refer to the journal's Authorship policy in the Editorial Policies and Ethical Considerations section for details on author listing eligibility. When entering the author names into Editorial Manager, the corresponding author will be asked to provide a CRediT contributor role to classify the role that each author played in creating the manuscript. Please see the **Project CRediT** website for a list of roles.

Abstract

Please provide an abstract which gives a concise statement of the intention, results or conclusions of the article. The abstract should not include any sub-headings.

- Abstracts for Research Papers should not exceed 250 words.
- Abstracts for theoretical or review articles should not exceed 250 words.
- Abstracts for brief communications should not exceed 80 words.

Keywords

Please provide appropriate keywords.

Acknowledgments

Contributions from anyone who does not meet the criteria for authorship should be listed, with permission from the contributor, in an Acknowledgments section. Financial and material support should also be mentioned. Thanks to anonymous reviewers are not appropriate.

Main Text File

As papers are double-blind peer reviewed, the main text file should not include any information that might identify the authors.

The main text file should be presented in the following order:

- Title
- Main text
- References
- Tables and figures (each complete with title and footnotes)
- Appendices (if relevant)

Supporting information should be supplied as separate files. Tables and figures can be included at the end of the main document or attached as separate files but they must be mentioned in the text.

- As papers are double-blind peer reviewed, the main text file should not include any information that might identify the authors. Please do not mention the authors' names or affiliations and always refer to any previous work in the third person.
- The journal uses British/US spelling; however, authors may submit using either option, as spelling of accepted papers is converted during the production process.

References

References in published papers are formatted according to the Publication Manual of the American Psychological Association (6th edition). However, references may be submitted in any style or format, as long as it is consistent throughout the manuscript.

Tables

Tables should be self-contained and complement, not duplicate, information contained in the text. They should be supplied as editable files, not pasted as images. Legends should be concise but comprehensive – the table, legend, and footnotes must be understandable without reference to the text. All abbreviations must be defined in footnotes. Footnote symbols: †, ‡, §, ¶, should be used (in that order) and *, **, *** should be reserved for P-values. Statistical measures such as SD or SEM should be identified in the headings.

Figures

Although authors are encouraged to send the highest-quality figures possible, for peer-review purposes, a wide variety of formats, sizes, and resolutions are accepted.

Click here for the basic figure requirements for figures submitted with manuscripts for initial peer review, as well as the more detailed post-acceptance figure requirements.

Legends should be concise but comprehensive – the figure and its legend must be understandable without reference to the text. Include definitions of any symbols used and define/explain all abbreviations and units of measurement.

Colour figures. At the editors' discretion, colour figures can be provided for use in the journal. Good quality photographs will be considered where they add substantially to the argument, to a maximum of three per article. These can be supplied electronically as TIF files scanned to at least 300dpi. If they are printed in colour, then they can be reproduced in colour online and black and white in print.

Supporting Information

Supporting information is information that is not essential to the article, but provides greater depth and background. It is hosted online and appears without editing or typesetting. It may include tables, figures, videos, datasets, etc.

Click here for Wiley's FAQs on supporting information.

Note: if data, scripts, or other artefacts used to generate the analyses presented in the paper are available via a publicly available data repository, authors should include a reference to the location of the material within their paper.

General Style Points

For guidelines on editorial style, please consult the **APA Publication Manual** published by the American Psychological Association. The following points provide general advice on formatting and style.

- **Language:** Authors must avoid the use of sexist or any other discriminatory language.
- **Abbreviations:** In general, terms should not be abbreviated unless they are used repeatedly and the abbreviation is helpful to the reader. Initially, use the word in full, followed by the abbreviation in parentheses. Thereafter use the abbreviation only.
- **Units of measurement:** Measurements should be given in SI or SI-derived units. Visit the Bureau International des Poids et Mesures (BIPM) website for more information about SI units.
- **Effect size:** In normal circumstances, effect size should be incorporated.
- **Numbers:** numbers under 10 are spelt out, except for: measurements with a unit (8mmol/l); age (6 weeks old), or lists with other numbers (11 dogs, 9 cats, 4 gerbils).

Appendix 1.2: Systematic review search strategy in EMBASE

Components of search:

1. Physical activity (exposure)
2. Cognitive function (outcome)
3. Longitudinal (type of evidence)

Database(s): Embase 1947-Present, updated daily

Search Strategy:

#	Searches
1	physical activity/ or fitness/ or exercise/
2	accelerometer/ or accelerometry/ or actimetry/
3	(exercis* or physical activit* or fitness* or acceleromet* or actigraph*).tw.
4	1 or 2 or 3
5	attention/ or executive function/ or memory/ or mental capacity/ or mental performance/ or intelligence/ or intelligence test/ or neuropsychological test/ or cognitive function test/ or alzheimer disease/ or dementia/ or cognitive defect/ep, et, pc
6	(cognition or cognitive d#sfunction or intelligence or intelligence tests or neuro?psychological tests or memory or dementia* or alzheimer* disease or IQ or memory or processing speed or visuospatial or executive function or cognitive decline or cognitive ability or cognitive impairment).tw.
7	5 or 6
8	cohort analysis/ or prospective study/ or longitudinal study/
9	(cohort or follow?up or longitudinal or prospective).tw.
10	8 or 9
11	4 and 7 and 10
12	limit 11 to (english language and yr="2005 - 2021")

Appendix 1.3: Search strategy test of sensitivity results

Paper	Embase	Medline	Psychinfo	Overall
Bowen, M.E. (2012). A prospective examination of the relationship between physi- cal activity and dementia risk in later life. American Journal of Health Promotion, 26, 333-340. doi:10.4278/ajhp.110311- QUAN-115	y	y	y	y
Chang, M., Jonsson, P.V., Snaedal, J., Bjornsson, S., Saczynski, J.S., Aspelund, T.,...Launer, L.J. (2010). The effect of midlife physical activity on cognitive function among older adults: AGES—Reykjavik Study. Jour- nals of Gerontology. Series A, Biological Sci- ences and Medical Sciences, 65, 1369-1374. doi:10.1093/gerona/glq152	y	y	y	y
de Bruijn, R.F., Schrijvers, E.M., de Groot, K.A., Witteman, J.C., Hofman, A., Franco, O.H.,...Ikram, M.A. (2013). The association between physical activity and dementia in an elderly population: The Rotterdam Study. European Journal of Epide- miology, 28, 277-283. doi:10.1007/s10654- 013-9773-3	y	y	n	y
Luck, T., Riedel-Heller, S.G., Luppa, M., Wiese, B., Kohler, M., Jessen, F.,...Maier, W. (2014). Apolipoprotein E epsilon 4 genotype and a physically active lifestyle in late life: Analysis of gene-environment interaction for the risk of dementia and Alzheimer's disease dementia. Psychologi- cal Medicine, 44, 1319-1329. doi:10.1017/ S0033291713001918	y	y	y	y
Heser, K., Wagner, M., Wiese, B., Prokein, J., Ernst, A., Konig, H.H.,...Eisele, M. (2014). Associations between dementia outcomes and depressive symptoms, leisure activities, and social support. Dementia and Geriat- ric Cognitive Disorders Extra, 4, 481-493. doi:10.1159/000368189	n	y	n	y
Podewils, L.J., Guallar, E., Kuller, L.H., Fried, L.P., Lopez, O.L., Carlson, M., & Lyketsos, C.G. (2005). Physical activity, APOE genotype, and dementia risk: Findings from the Cardiovascular Health Cognition	y	y	n	y
Rovio, S., Kareholt, I., Helkala, E.L., Viitanen, M., Winblad, B., Tuomilehto, J.,... Kivipelto, M. (2005). Leisuretime physi- cal activity at midlife and the risk of de- mentia and Alzheimer's disease. Lancet Neurology,	y	y	n	y
Sabia, S., Dugravot, A., Dartigues, J.F., Abell, J., Elbaz, A., Kivimaki, M., & Singh-Manoux, A. (2017). Physical activity, cognitive decline, and risk of dementia: 28 year follow- up of Whitehall II cohort study. BMJ, 357, j2709. doi:10.1136/bmj.j2709	y	y	y	y
Tolppanen, A.M., Solomon, A., Kulmala, J., Kareholt, I., Ngandu, T., Rusanen, M.,... Kivipelto, M. (2015). Leisure-time physi- cal activity from mid- to late life, body mass index, and risk of dementia. Alzheimer's & Dementia, 11, 434-443. doi:10.1016/j. jalz.2014.01.008	y	y	y	y
Baezner, H., Blahak, C., Poggesi, A.,... Inzitari, D. (2012). Physical activity prevents progression for cognitive impairment and vascular dementia: Results from the LADIS (Leukoaraiosis and Disability) study. Stroke, 43, 3331-3335. doi:10.1161/ STROKEAHA.112.661793	y	y	n	y
Fenesi, B., Fang, H., Kovacevic, A., Oremus, M., Raina, P., & Heisz, J.J. (2017). Physical exercise moderates the relationship of apolipoprotein E (APOE) genotype and dementia risk: A population-based study. Journal of Alzheimer's Disease, 56, 297-303. doi:10.3233/JAD-160424	n	n	n	n

Appendix 1.4: Screening eligibility checklist*

**Amendments made between versions 1 and 2 are shown in stage 3 (full-text) section in bold red text.*

Systematic Review Article Screening Tool (v2)

Developed using:

- Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. K. (2019). Best practice guidelines for Abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods*, *10*(3), 330-342.

Stage 1: Citation, title and abstract screening

Citation, Title, and Abstract Screening

1. Does the **citation** indicate publication on or after 2005?
    a. **Yes:** continue screening
    b. **No**: stop screening
2. Does the **title or abstract** use English?
    a. **Yes**: continue screening
    b. **No:** stop screening
3. Does the **title or abstract** indicate that this is **NOT** a review paper (systematic review/meta-analysis) or an intervention study, e.g. an RCT.
    a. **Yes**: continue screening
    b. **No:** stop screening
4. Does the **title or abstract** indicate that study population is **NOT** children/adolescents
    a. **Yes**: continue screening
    b. **No:** stop screening
5) Does the **title or abstract** indicate that the participants were free of cognitive impairment at baseline?
    a. **Yes**: continue screening
    b. **No:** stop screening
        - For example, the title or abstract indicates that participants had dementia diagnosis or MCI at baseline

Stage 2: Abstract screening

6. Does the **abstract** indicate that the age of the sample was **≥45 - <80** at baseline?
    a. **Yes or unsure/unclear**: continue screening
    b. **No:** stop screening
        - For example if abstract indicated that entire sample were <45, or >80 at baseline, or that the mean age of the sample at baseline was out with ≥45 -**<80** at baseline
7. Does the **abstract** indicate that this study was longitudinal in design?
    a. **Yes or unsure/unclear**: continue screening
        - Key words: prospective, cohort study, follow-up, multiple time-points, waves
    b. **No:** stop screening
        - For example study described as cross-sectional, retrospective, or as an RCT or other type of intervention study
8. Does the **abstract** indicate that physical activity was studied as primary exposure of interest?
    a. **Yes or unsure/unclear**: continue screening

- Key words: physical activity, exercise, fitness, accelerometer, walking. If the paper mentions PA as a 'joint' primary exposure, e.g. 'Cognitive and physical activity', include the paper.

b. **No:** stop screening

- For example, if PA is not included at all, or if PA is mentioned only as a covariate; or if the paper is looking more generally at range of predictors of cognitive function

9. Does the abstract indicate that cognitive function/decline, including dementia or Alzheimer's diagnosis, using neuropsychological test(s) was measured?

a. **Yes or unsure/unclear**: continue screening

b. **No:** stop screening

- For example, if the outcome is Dementia/Alzheimer diagnosis according to linked health records (diagnosis), but without any report of neuropsychological tests.

Decision: **Should this article be included for full-test screening?**

a. **Yes**, all X screening questions answered Yes or Unclear

b. **No**, at least one answers definitely "No"

Stage 3: Full-text screening:

All of the criteria above still apply. Although they should have been assessed at the ti/ab stage is is worth re-checking the key info such as participant age range and study design.

Inclusion Criteria:

Design of study:

Include if:

Longitudinal design (exclude cross-sectional).

If you are unsure then a quick way can be to search within the document for 'longitudinal' or 'cross-sectional'.

Exclude if:

Any other type of design, e.g. intervention or retrospective.

Cognitive function:

Include if:

- Measured by use of objective tests (i.e. neuropsychological measures) including screening measures such as the MMSE **and phone interview measures such as the TICS.**
- The study describes a process whereby an outcome (e.g. dementia) is determined by the use of a screening measure followed by clinical assessment – e.g. "further assessment subject to DSM diagnostic criteria",

Exclude if:

- If the measures is subjective: i.e. subjective memory complaint, **or subjective questions about change in cognition**
- If study only reports dementia diagnosis, without reference to cognitive measures (e.g. linked to insurance records or based on clinical interview only)

Physical activity (many studies look at a range of predictors without any clear focus on PA)

Include if:

- Measured either objectively (accelerometer), or using self-report questionnaires (either validated measures or simple questions about activity levels)

	~~- PA is the main exposure or, or one of two exposures (e.g. looking at cognitive activity and PA, but not a whole range of exposures)~~ - **PA is the main exposure of interest and not merely included as a range of predictors of CF** - **If PA is one of two main exposures (e.g. PA & cognitive activity, or PA & APOE-e4). But not more than two.**
Exclude if:	- There is no measure of physical activity - PA is included as one of multiple exposures without clear focus

Appendix 1.5: Data extraction form

Study	Country	Age at baseline	N at baseline	Follow-up time	Exposure measure	Exposure classification	Outcome measure	Impairment classification	Key findings	Covariates included	Modelling and adjustment strategies
(Barha et al., 2020)											
(Etgen et al., 2010)											
(Fiocco et al., 2012)											
(Gow et al., 2017)											
(Halloway et al., 2017)											
(Hamer et al., 2018)											
(Iso-Markku et al., 2016)											
(Larson et al., 2006)											
(Morgan et al., 2012)											
(Ogino et al., 2019)											
(Papenberg et al., 2016)											
(Podewils et al., 2005)											
(Rabin et al., 2019)											
(Rajan et al., 2015)											
(Reas et al., 2019)											
(Rovio et al., 2005)											
(Sabia et al., 2017)											
(Tan et al., 2017)											
(Thibeau et al., 2016)											
(Verghese et al., 2009)											
(Willey et al., 2016)											

Appendix 1.6: Risk of bias/quality ratings

Yes	No	Unclear	Not applicable

Study	JBI Critical Appraisal Checklist for Cohort Studies											CASP Cohort study checklist
	1	2	3	4	5	6	7	8	9	10	11	A2
	Were the two groups similar and recruited from the same population?	*Were the exposures measured similarly to assign people to both exposed & unexposed groups ?*	*Was the exposure measured in a valid & reliable way ?*	*Were confounding factors identified?*	*Were strategies to deal with confounding factors stated?*	*Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure?)*	*Were the outcomes measured in a valid and reliable way?*	*Was the follow-up time reported and long enough for outcomes to occur?*	*Was follow-up complete, and if not, were the reasons to loss to follow up explored?*	*Were strategies to address incomplete follow-up utilised?*	*Was appropriate statistical analysis used?*	*Was the cohort recruited in an acceptable way?*
(Barha et al., 2020)	u	y	n	y	y	y	y	y	n	n	y	u
(Etgen et al., 2010)	n	y	n	y	y	y	y	n	y	n/a	y	n
(Fiocco et al., 2012)	n	y	y	y	y	y	y	n	u	n/a	y	u
(Gow, Pattie, & Deary, 2017)	u	y	n	y	y	u	y	y	n	n	y	n
(Halloway, Wilbur, Schoeny, & Barnes, 2017)	u	y	y	y	y	y	y	y	n	n	y	n

Study	JBI Critical Appraisal Checklist for Cohort Studies											CASP Cohort study checklist
	1	2	3	4	5	6	7	8	9	10	11	A2
	Were the two groups similar and recruited from the same population?	*Were the exposures measured similarly to assign people to both exposed & unexposed groups ?*	*Was the exposure measured in a valid & reliable way ?*	*Were confounding factors identified?*	*Were strategies to deal with confounding factors stated?*	*Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure?)*	*Were the outcomes measured in a valid and reliable way?*	*Was the follow-up time reported and long enough for outcomes to occur?*	*Was follow-up complete, and if not, were the reasons to loss to follow up explored?*	*Were strategies to address incomplete follow-up utilised?*	*Was appropriate statistical analysis used?*	*Was the cohort recruited in an acceptable way?*
(Hamer, Terrera, & Demakakos, 2018)	n	y	n	y	y	y	n	y	y	n/a	y	y
(Iso-Markku et al., 2018)	n	y	n	y	y	u	y	y	n	n	y	u
(Larson et al., 2006)	y	y	n	y	y	y	y	y	u	y	y	y
(Morgan et al., 2012)	n	y	n	y	y	u	y	y	n	y	y	n
(Ogino, Manly, Schupf, Mayeux, & Gu, 2019)	u	y	y	y	y	y	y	y	n	n	y	n
(Papenberg et al., 2016)	n	y	y	y	y	y	y	y	u	n/a	y	u
(Podewils et al., 2005)	n	y	y	y	y	y	y	y	n	n	y	n
(Rabin et al., 2019)	u	y	y	y	y	y	y	y	n	n	y	u
(Rajan et al., 2015)	u	y	n	y	y	u	y	y	u	n/a	y	n

Study	JBI Critical Appraisal Checklist for Cohort Studies											CASP Cohort study checklist
	1	2	3	4	5	6	7	8	9	10	11	A2
	Were the two groups similar and recruited from the same population?	*Were the exposures measured similarly to assign people to both exposed & unexposed groups ?*	*Was the exposure measured in a valid & reliable way ?*	*Were confounding factors identified?*	*Were strategies to deal with confounding factors stated?*	*Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure?)*	*Were the outcomes measured in a valid and reliable way?*	*Was the follow-up time reported and long enough for outcomes to occur?*	*Was follow-up complete, and if not, were the reasons to loss to follow up explored?*	*Were strategies to address incomplete follow-up utilised?*	*Was appropriate statistical analysis used?*	*Was the cohort recruited in an acceptable way?*
(Reas, Laughlin, Bergstrom, Kritz-Silverstein, & McEvoy, 2019)	n	y	n	y	y	y	y	y	n	y	y	u
(Rovio et al., 2005)	y	y	n	y	y	u	y	y	y	n	y	u
(Sabia et al., 2017)	n	y	y	y	y	u	y	y	y	y	y	n
(Tan et al., 2017)	u	y	n	y	y	y	u	y	u	n/a	y	u
(Thibeau, McFall, Wiebe, Anstey, & Dixon, 2016)	u	y	y	y	y	y	y	y	n	y	y	u
(Verghese, Wang, Katz, Sanders, &	u	y	n	y	y	y	y	y	n	n	y	n

Study	JBI Critical Appraisal Checklist for Cohort Studies											CASP Cohort study checklist
	1	2	3	4	5	6	7	8	9	10	11	A2
	Were the two groups similar and recruited from the same population?	*Were the exposures measured similarly to assign people to both exposed & unexposed groups ?*	*Was the exposure measured in a valid & reliable way ?*	*Were confounding factors identified?*	*Were strategies to deal with confounding factors stated?*	*Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure?)*	*Were the outcomes measured in a valid and reliable way?*	*Was the follow-up time reported and long enough for outcomes to occur?*	*Was follow-up complete, and if not, were the reasons to loss to follow up explored?*	*Were strategies to address incomplete follow-up utilised?*	*Was appropriate statistical analysis used?*	*Was the cohort recruited in an acceptable way?*
Lipton, 2009)												
(Willey et al., 2016)	n	y	n	y	y	y	y	y	n	n	y	n

Appendix 2.1: NHS Research Ethics Committee approval

NHS
Health Research Authority

North West - Haydock Research Ethics Committee
3rd Floor - Barlow House
4 Minshull Street
Manchester
M1 3DZ

Telephone: 02071048103

29 June 2021

Professor Naomi Allen
UK Biobank Limited
Clinical Trial Service Unit and Epidemiological Studies Unit
Nuffield Department of Population Health, The Big Data Institute
University of Oxford, Oxford
OX3 7LF

Dear Professor Allen,

Title of the Database:	**UK Biobank: a large scale prospective epidemiological resource**
Designated Individual:	**Mrs Samantha Welsh**
REC reference:	**21/NW/0157**
IRAS project ID:	**299116**

Thank you for responding to the Favourable Opinion with Condition. I can confirm the REC has received the documents listed below and that these comply with the approval conditions detailed in our correspondence dated 18 June 2021.

Documents received

The documents received were as follows:

Document	*Version*	*Date*
Other [Participant Withdrawal Form]	3.0	25 June 2021

Approved documents

The final list of approved documentation for the study is therefore as follows:

Document	*Version*	*Date*
Human Tissue Authority licence [HTA Licences]	26 Feb 2020	26 February 2020
IRAS Checklist XML [Checklist_30042021]		30 April 2021
IRAS Checklist XML [Checklist_29062021]		29 June 2021
Other [EGF]	3.0	29 October 2007

Other [Data Dictionary Showcase]	1	29 April 2021
Other [Annual Progress Report Form]	1.0	29 April 2021
Other [UK Biobank Protocol]	21 March 2007	21 March 2007
Other [Protocol Addendum 1]	09 April 2009	09 April 2009
Other [Protocol Addendum 2]	2 July 2009	02 July 2009
Other [Table 1 Samples Collected and Stored at UKB]	1.0	29 April 2021
Other [CV NAllen]	April 2021	29 April 2021
Other [BGMini]	1.0	17 July 2019
Other [Participant Feedback Survey (Cardiac Monitoring)]	1.0	06 November 2020
Other [In Clinic Application PIL FAQ (Cardiac Monitoring)]	2.3	12 January 2021
Other [Self Application PIL & FAQ (Cardiac Monitoring)]	3.1	06 November 2020
Other [Reminder to Return UKB Heart Monitor 30 days]	2.1	26 April 2021
Other [Reminder to Return UKB Heart Monitor 58 days]	1.1	26 April 2021
Other [Reminder to start wearing the UKB heart monitor]	1.1	26 April 2021
Other [Thank you for participating in UKB Heart Monitor]	1.1	26 April 2021
Other [Thank you for returning your UKB Heart Monitor]	1.1	26 April 2021
Other [Cog Funct invitation email]	1.1	12 February 2021
Other [Cog Funct non responders reminder email]	1.1	27 April 2021
Other [Cog Funct reminder partial responders email]	1.1	27 April 2021
Other [Cog Funct Last Chance reminder email]	1.2	27 April 2021
Other [Pre_Imaging Visit Questionnaire]	2.4	27 November 2020
Other [Cheadle COVID-19 Repeat Imaging Invite]	1.1	01 December 2020
Other [COVID-19 Info Leaflet Second Assessment]	1.5	17 December 2020
Other [Stockport COVID-19 Repeat Imaging Postal Appt Confirmation]	1.4	27 November 2020
Other [Cheadle COVID-19 Repeat Imaging Email Appt Confirmation]	1.1	01 December 2020
Other [Cheadle COVID-19 Repeat Imaging appointment reminder SMS]	1.0	20 October 2020
Other [Cheadle COVID-19 Repeat Imaging 1st Reminder]	1.1	01 December 2020
Other [Cheadle COVID-19 Rpt Imaging 2nd Reminder]	1.0	20 October 2020
Other [Cheadle Repeat Imaging 2nd Invite]	1.2	17 December 2020
Other [COVID Secure Measures for Participants]	1.3	17 December 2020
Other [DFP UKB Invitation email]	1.1	13 June 2019
Other [Website content UKB (DFP)]	1.1	13 June 2019
Other [DFP UKB Participant Study Summary]	1.0	01 April 2019
Other [food pref invitation email]	1.1	27 April 2021
Other [food pref reminder non responder reminder email]	1.1	27 April 2021
Other [food pref reminder partial responders email]	1.1	17 April 2021
Other [food pref last chance reminder email]	1.1	27 April 2021
Other [Food preferences Metadata]	1.1	20 May 2019
Other [Email Appt Confirmation]	1.1	28 April 2021

Other [IF Letter to GP and Specialist Report]	3.0	29 April 2021
Other [IF Letter to Participant]	3.0	29 April 2021
Other [Imaging Appointment Confirmation Letter]	4.0	28 April 2021
Other [Imaging Invitation Stockport]	3.0	26 April 2021
Other [Imaging Invitation Postal]	2.0	23 January 2020
Other [Info Leaflet]	11	29 September 2019
Other [Second, Third & Fourth Invitation to attend Imaging]	0.2	26 April 2021
Other [Imaging 2nd reminder]	1.0	27 April 2021
Other [Imaging SMS Appointment Reminder]	1.0	26 April 2021
Other [Reminder postal invitation (Imaging)]	1.0	17 January 2020
Other [Pain Invitation email]	1.1	27 April 2021
Other [Pain non responders reminder email]	1.1	27 April 2021
Other [Pain partial responders reminder]	1.1	27 April 2021
Other [Pain Last Chance reminder email]	1.1	27 April 2021
Other [Pain questionnaire Metadata]	2.4	07 November 2018
Other [First reminder Repeat Imaging Stockport]	22 April 2021	22 April 2021
Other [IF Letter to GP and Specialist Report]	3.0	29 April 2021
Other [IF Letter to Participant]	3.0	29 April 2021
Other [Imaging Appointment Confirmation Letter]	4.0	28 April 2021
Other [Invitation Repeat Imaging Invite]	22 April 2021	22 April 2021
Other [Repeat Imaging appointment confirmation email]	22 April 2021	22 April 2021
Other [Second Reminder Repeat Imaging Stockport]	22 April 2021	22 April 2021
Other [Invitation Postal (Repeat Imaging)]	05 Feb 2019	05 February 2019
Other [2nd Invitation Repeat Imaging Stockport]	28 April 2021	28 April 2021
Other [Imaging SMS Appointment Reminder]	1.0	26 April 2021
Other [Sleep Metadata Questionnaire]	2.0	12 March 2020
Other [Sleep Questionnaire Invitation Email]	0.1	26 July 2019
Other [Sleep Questionnaire Non Responders]	0.1	26 July 2019
Other [Sleep Questionnaire Partial Responders]	0.1	26 July 2019
Other [Sleep Questionnaire Final Reminder]	0.1	26 July 2019
Other [SMS Contact Details Update]	29 April 2021	29 April 2021
Other [Touch Screen Questionnaire]	29 April 2021	29 April 2021
Other [Annual Participant Newsletter]	27 November 2020	27 November 2020
Other [Document 1 Protocol Extract]	21 Sept 2006	21 September 2006
Other [Table 2 Longitudinal Health Outcomes Data]	30 April 2021	30 April 2021
Other [Table 3 Future Plans]	30 April 2021	30 April 2021
Other [Touch Screen Questionnaire Reaction Time Test (Snap)]	1.3	05 April 2018
Other [Touch Screen Questionnaire Picture Vocabulary Test]	0.1	09 November 2017
Other [Touch Screen Questionnaire Fluid Intelligence Test]	1.3	09 November 2017
Other [Touch Screen Questionnaire Pairs Test]	1.3	09 November 2017

Other [Touch Screen Questionnaire Numeric Memory Test]	1.2	09 November 2017
Other [Touch Screen Questionnaire Trail Making Test]	0.1	09 November 2017
Other [Touch Screen Questionnaire Matrix Fattern Completion]	0.1	09 November 2017
Other [Touch Screen Questionnaire Prospective Memory Test]	1.2	09 November 2017
Other [Participant Withdrawal Form]	3.0	25 June 2021
Participant consent form [20061124 Consent form]	20061124	24 November 2006
Participant consent form [Consent Form Imaging]	29 Jan 2014	29 January 2014
Participant consent form [DFP Informed Consent Form]	4.0	01 April 2019
Participant consent form [DFP Study Partner Informed Consent Form]	2.0	01 April 2019
Participant consent form [DFP Informed Consent Form Mezurio]	4.0	01 April 2019
Participant information sheet (PIS) [Participant Information Leaflet]	21 April 2010	21 April 2010
Participant information sheet (PIS) [Repeat Imaging Info Leaflet]	Sept 2019	29 September 2019
Participant information sheet (PIS) [DFP UKB Participant Information Sheet]	1.2	21 June 2019
Participant information sheet (PIS) [DFP Study Partner Information Sheet]	2.0	01 April 2019
Participant information sheet (PIS) [DFP Participant Information Sheet Mezurio]	4.1	21 June 2019
Participant information sheet (PIS) [COVID-19 Info Leaflet Second Assessment]	1.4	09 December 2020
Protocol for management of the tissue bank [Access procedures]	1.0	29 November 2011
REC Application Form [RTB_Form_30042021]		30 April 2021
Response to Additional Conditions Met		29 June 2021

IRAS project ID299116	**Please quote this number on all correspondence**

Yours sincerely,

[redacted]

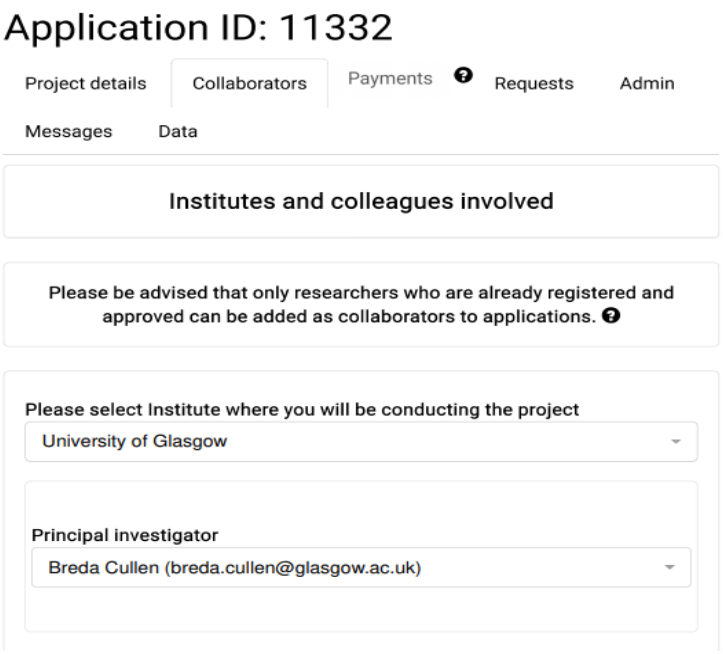
Miss Yasmin King (ANutr)
Approvals Officer

E-mail: haydock.rec@hra.nhs.uk

Copy to: *Mrs Samantha Welsh, UK Biobank Limited*

Appendix 2.2: Approved UK Biobank research application number 11332

Page 1/2



Page 2/2

Appendix 2.3: NHS Lanarkshire acknowledgement of research

FW: Notification of Biobank research

CT Campbell, Thomas - Trainee Clinical Psychologist <Thomas.Campbell@lanarkshire.scot.nhs.uk>
To Thomas Campbell (PGR)

Reply | Reply All | Forward | ...
Wed 16/09/2020 08:11

11332collaborators.pdf 247 KB
20150330 11332 MTA B C executed.pdf 1 MB

To: Campbell, Thomas - Trainee Clinical Psychologist <Thomas.Campbell@lanarkshire.scot.nhs.uk>
Cc: McGonigal, Elizabeth (MK) Senior R&D Facilitator <Elizabeth.Mcgonigal@lanarkshire.scot.nhs.uk>; Dolier, Cynthia - Lead R&D Facilitator <Cynthia.Dolier@lanarkshire.scot.nhs.uk>; Quinn, Lorraine (MK) - Senior R&D Facilitator <Lorraine.Quinn@lanarkshire.scot.nhs.uk>
Subject: FW: Notification of Biobank research

Hi Thomas – interesting thesis. There is no recruitment or use of NHS Lanarkshire facilities involved, so I don't think we need to carry out a formal review (*we normally 'acknowledge' rather than 'approve' biobank studies*).

Am I right to assume that you / your supervisor has to make a formal project application for access to UK Biobank data for this specific project (*or does the "contract" you mention in your proposal allow your supervisor free access?*) – what's the process around that / do they issue an approval letter of similar? If so, can we get a copy of that approval?

Raymond

Raymond Hamill
Senior R&D Manager
NHS Lanarkshire
Monklands Hospital
Monkscourt Avenue
Airdrie ML6 0JS

e: raymond.hamill@lanarkshire.scot.nhs.uk
t: 01236712446
m: 07779161388

From: Campbell, Thomas - Trainee Clinical Psychologist
Sent: 04 September 2020 16:07
To: Hamill, Raymond (NHS Lanarkshire) - Senior R&D Manager
Cc: breda.cullen@glasgow.ac.uk
Subject: Notification of Biobank research

Dear Raymond,

I hope I am writing to the appropriate person regarding this request.

I am a Trainee Clinical Psychologist, on the University of Glasgow programme, with my research supervised by Dr Breda Cullen (c.c.).

For my thesis I am conducting research out with NHSL, using the UK Biobank Database, which requires me to inform NHSL R&D of this activity.

Attached is a copy of my university approved project proposal, as well as the ethics letters pertaining to UK Biobank.

I'd be most grateful if you could acknowledge this notification of research activity, or direct me to the appropriate person within our R&D department.

Kind regards, Thomas Campbell

Appendix 2.4: Matching concepts to UK Biobank data process

Concept	Biobank match	Link
Acculturation	Year moved to UK	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=3659
Adiposity	BMI	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21001
	Waist circumference	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=48
Adverse experiences	Traumatic events	https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=145
Age	Age in years	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=21003
Alcohol	Frequency	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1558
	Binge	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20416
All medical diagnoses	First occurrences (pre baseline)	https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=1712
Brain health	Grey matter normalized	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=25005
	White matter grade	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25781
	(tract integrity) DTI	https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=135
	Hippocampus volume	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=26562 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=26593
	Volume of anterior cingulate cortex	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25838 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25839
	Volume of dlPFC (represented by superior and middle	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25786 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25787

	frontal gyrus measures from Freesurfer)	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25788 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25789
Calcium	Blood biochemistry	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=30680
Cardiovasular risk	Atrial fibrillation Arterial stiffness Blood pressure	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=131350 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21021 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4079 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4080
Childhood IQ	Vocab measure from imaging visit	
Childhood PA	Not measured	
Cholesterol	Blood biochemistry HDL	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=30690 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=30690
Cognitive function	Baseline and follow up	https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100026 https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=116
Diet		
Education	Degree or not	Derived dichotomous variable representing whether participant has college degree or not.
Ethnicity	Ethnic category	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=21000
Family HX dementia/PD	Parent or sibling with diagnosis	
Genetic risk	Dementia risk	Derived variables: described at https://choishingwan.github.io/PRS-Tutorial/ldpred/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4596916/
Inflammation	C-reactive protein	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=30710
Living alone	Number in household	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=709
Manual work	Job involves manual /heavy work	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=816

Marital status	How are people in house related to participant	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6141
Maternal smoking	Maternal smoking around birth	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=1787
Physical activity	Baseline (self-report) MET scores (accelerometer)	https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=54 https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=1009
Pollution	Inverse distance to major road	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24012
Psychotropic and anti hypertensive Medication	Medication touchscreen questions Psychotropics from verbal interview; cardiometabolic from touchscreen multiple choice	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20003
SES	Household Income Townsend deprivation index	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=738 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=189
Sex	Sex baseline	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=31
Smoking status	Tobacco history	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24012
Social network	Frequency of friend/family visits	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1031
Sodium	Salt added to food	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1478
Trait anxiety	Neuroticism score Are you a worrier	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20127

Appendix 2.5: Description of covariate measures within UK Biobank

Genetic

A polygenic risk score for a combination of dementia genes was standardised as a z-score and treated as a continuous variable. A variable reflecting APOE genotype (number of e4 alleles) was treated as continuous. Technical covariates related to genetic variables were also included in adjusted models.

Sociodemographic variables

Age was calculated by subtracting date of birth from assessment date. Acculturation was derived by subtracting date years in UK from date of birth. Gender, ethnic group and country of origin were self-reported via touchscreen interview. Townsend deprivation indices were derived from postcode of residence (categorised into quintiles with one representing the least, and five the most, deprived). Income was treated as an ordinal variable reflecting which category (low to high) participants belonged to. Education was dichotomised to reflect whether participant had a university/college degree or not. A variable representing exposure to pollution was derived using the inverse distance from a person's home address to a major road and split into quintiles with one representing the farthest distance and five the nearest. Living alone was recorded by touch-screen interview and treated as dichotomous (yes/no).

Health behaviours, medical risk factors and physical measurements

Past and current smoking habits were self-reported via touch-screen interview and a binary variable (ever a smoker) was derived. Sodium intake and alcohol frequency were also self-reported and treated as ordinal variables. Frequency of binge drinking (defined as greater than 6 units in one sitting) was recorded later using the web-questionnaire treated as an ordinal variable. A range of key biochemistry markers including and were measured using blood tests. Cholesterol (hdl, ldl and total) and calcium are treated as continuous variables expressed in mmol/L, and inflammation (serum C-reactive protein level, expressed in mg/L) is treated as ordinal. BMI was calculated by kg/m^2 and treated as a continuous variable, and waist circumference was measured in cm and treated as continuous. Systolic and diastolic blood pressure were measured, expressed in mmHG, and treated as continuous variables.

Medical comorbidities and medications

Relevant medical diagnoses indicated by the conceptual model were all recorded using linked health records according to ICD-11 codes. Dichotomous variables whereby 'yes' represented the participant receiving a diagnosis on or before the date of baseline assessment. The relevant diagnostic categories were: neurological conditions, dementia, diabetes, chronic kidney disease, head injury, cerebrovascular disease, chronic lung disease, musculoskeletal conditions, mood disorder, psychotic illness and cardiovascular disease.

Dichotomous variables for antihypertensive and psychotropic medications were derived, coded 'yes' if the participant self-reported being on any medication within these categories.

Mental health variables

Trait anxiety was represented by a question asked at baseline ("are you a worrier?") and treated as a dichotomous variable (yes/no). Neuroticism was measured using 12 self-report questions at baseline (e.g., "Are you an irritable person?) and a continuous variable was derived representing the number of items the participant answered yes to. Experience of trauma/adverse experiences was measured using the web-based questionnaire (e.g., "When I was growing up, someone sexually molested me", and a dichotomous variable representing any incidence of trauma was derived. Mood was measured using four questions regarding depression symptoms, which were summed and treated as a continuous variable.

Other

Follow-up duration was calculated by subtracting the date of follow-up assessment from the date of baseline assessment and expressed in years.

Covariates included in adjusted models:

Alcohol binge, alcohol frequency, antihypertensive medication, apoe-e4 allele count, BMI, cardiovascular disease dx, dementia genetic risk score, diabetes dx, distance to major road, friend and family visits, follow-up time, gender, hdl cholesterol, head injury dx, household income, kidney disease dx, KJ of energy, ldl cholesterol, living alone status, manual work, mood disorder dx, musculoskeletal dx, neurological disorder dx, neuroticism score, psychosis dx, psychotropic medication, salt added to food, smoking status, technical genetic covariates, Townsend deprivation score, trauma status, waist circumference, worrier status. Also adjusted for technical covariates used with genetic risk scores

Appendix 2.6: Sensitivity analysis results table

Cross sectional models

Exposure	Cognitive score	Unadjusted, within the total sample					Unadjusted, within the sample that had complete covariate data				
		n	Estimate[a]	95% CI	P (uncorr)	p (FDR)	n	Estimate	95% CI	p (uncorr)	p (FDR)
Total PA, self-report (MET hrs/week)	Reaction Time	305,294	.000131	.0000659, .0001962	.0001	<.0001	29,810	.0001306	-.0001092 .0003704	.2857	.3429
	Pairs Matching	300,847	-.0005355	-.0006077, -.004634	<.0001	<.0001	29,664	-.0004247	-.0006967 -.0001528	.0022	.0033
	Reasoning	100,204	-.0023377	-.0024488, -.0022266	<.0001	<.0001	12,438	-.002656	-.002994 -.0023179	<.0001	<.0001
	Numeric Memory	31,854	-.0012581	-.014492, -.0010669	<.0001	<.0001	3,613	-.0013487	-.001991 -.0007064	<.0001	<.0001
	Global CF	300,915	-.0004846	-.0005331,-.0004362	<.0001	<.0001	29,695	-.0005834	-.0007574 -.0004093	<.0001	<.0001
	Prospective Memory	89,022	.9981991	.9979159, .998482	<.0001	<.0001	2,548	.9986968	.9955481, 1.001856	.4183	.4183

Abbreviations: CF, cognitive function; CI, confidence interval; FDR, False Discovery Rate; MET, Metabolic Equivalent of Task; PA , physical activity; Uncorr, uncorrected

a: All expressed as z scores, except Prospective Memory which is expressed as an odds ratio

b: Adjusted for: alcohol binge, alcohol frequency, antihypertensive medication, apoe-e4 allele count, BMI, cardiovascular disease dx, dementia genetic risk score, diabetes dx, distance to major road, friend and family visits, gender, hdl cholesterol, head injury dx, household income, kidney disease dx, KJ of energy, ldl cholesterol, living alone status, manual work, mood disorder dx, musculoskeletal dx, neurological disorder dx, neuroticism score psychosis dx, psychotropic medication, salt added to food, smoking status, Townsend deprivation score, trauma status, waist circumference, worrier status. Also adjusted for technical covariates used with genetic risk scores.

c: Probability adjusted using the Simes-Benjamini-Hochberg method implemented in the Stata qqvalue package.

d: Global CF = mean of z scores on four tests (assuming at least two non-missing values)

Longitudinal models

Exposure	Cognitive score	Unadjusted, within the total sample					Unadjusted, within the sample that had complete covariate data				
		n	**Estimate**	**95% CI**	***p* (uncorr)**	***p* (FDR)[b]**	***n***	**Estimate**	**95% CI**	***p* (uncorr)**	***p* (FDR)[b]**
Total PA, self-report (MET hrs/week)	Reaction time	30,153	-.0000553	-.0002922, .0001816	.6474	.6474	6,816	.0000752	-.0006065, .0004562	.7816	.7816
	Pairs matching	29,845	-.000428	-.0006911, -.0001649	.0014	.0015	6,780	-.0003613	-.0009561, .0002335	.2338	.2476
	Prospective Memory (	30,330	.9971918	.9965854 .9977987	<.0001	<.0001	6,840	. 9971716	.9955542, .9987915	.0006	.0007
	Reasoning	29,801	-.0027912	-.0030282, -.0025542	<.0001	<.0001	6,779	-.0032087	-.0037301, -.0026873	<.0001	<.0001
	Numeric Memory	22,321	-.001635	-.0019111, -.0013588	<.0001	<.0001	5,006	-.0021687	-.0028092, -.0015282	<.0001	<.0001
	Symbol Digit Substitution	21,831	-.0016325	-.0019087, -.0013564	<.0001	<.0001	4,886	-.0017565	-.0023865, -.0011266	<.0001	<.0001
	Paired Associate Learning	21,343	-.0015357	-.0017798, -.0012917	<.0001	<.0001	4,805	-.0014098	-.0019326, -.0008871	<.0001	<.0001
	Tower Rearranging	21,626	-.0011721	-.0014512, -.0008931	<.0001	<.0001	4,865	-.001141	-.0017697, -.0005123	.0004	.0005
	Matrix pattern completion	21,804	-.0019887	-.0022625, -.001715	<.0001	<.0001	4,881	-.0021046	-.0027152, -.0014941	<.0001	<.0001
	Trails A (time)	21,709	-.0013174	-.0015988, -.0010359	<.0001	<.0001	4,874	-.0015953	-.0022335, -.0009571	<.0001	<.0001
	Trails B (time)	21,225	- .0019304	-.0022155, -.0016453	<.0001	<.0001	4,827	- .0023499	-.0029889, -.0017108	<.0001	<.0001
	Trails B – A (time)	21,225	-.0015662	-.0018518, -.0012806	<.0001	<.0001	4,827	-.0019600	-.002588, -.0013319	<.0001	<.0001
	Trails B (errors)	21,698	-.002139	-.0026008, -.0016773	<.0001	<.0001	4,866	-.0027761	-.0037931, -.001759	<.0001	<.0001

	Processing Speed (comp)	21,767	-.0009248	-.0011378, -.0007117	<.0001	<.0001	4,873	-.000999	-.0014791, -.000519	<.0001	<.0001
	Executive Function (comp)	21,742	-.001489	-.0017003, -.0012777	<.0001	<.0001	4,871	-.0016902	-.002165, -.0012153	<.0001	<.0001
	Reasoning (comp)	21,655	-.0024415	-.0026714, -.0022116	<.0001	<.0001	4,866	-.002725	-.0032274, -.0022226	<.0001	<.0001
	Memory (comp)	22,467	-.0011893	-.0013674, -.0010112	<.0001	<.0001	5,029	-.0012647	-.0016617, -.0008677	<.0001	<.0001
	Global CF (comp)	30,048	-.0012926	-.0014324, -.0011529	<.0001	<.0001	6,804	-.0006555	-.0010004 -.0003107	.0002	.0003
Exposure	**Cognitive score**	**Unadjusted**					**Adjusted**[a]				
Physical Activity, accelerometery (milligravity units)		***n***	**Estimate**	**95% CI**	***p* (uncorr)**	***p* (FDR)**	***n***	**Estimate**	**95% CI**	***p* (uncorr)**	***p* (FDR)**
	Reaction time	14,307	.0029557	.0012409, .0046705	.0007	.0063	3,935	.0027609	-.0004749, .0059968	.1334	.8058
	Pairs matching	14,164	-.0009559	-.0028609, .0009492	.3254	.3584	3,919	-.0013041	-.0049577, .0023496	.7460	.8393
	Prospective Memory	14,392	1.006038	1.0007, 1.011405	.0266	.0798	3,828	1.001839	.9904113, 1.013399	.9162	.9162
	Reasoning	14,148	.0009323	-.0026586, .000794	.2898	.3478	3,919	-.000163	-.0033863, .0030603	.8516	.9017
	Numeric Memory	9,901	.0019117	-.0001713, .0039947	.0720	.1566	2,837	.0019519	-.0020293, .0059331	.3824	.8393
	Symbol Digit Substitution	9,683	.0030884	.0010321, .0051447	.0032	.0162	2,772	.0004748	-.0034015, .0043512	.6971	.8393
	Paired Associate Learning	9,522	.0015174	-.0003052, .0033401	.1027	.1849	2,731	.0013632	-.0019934, .0047198	.6143	.8393

	Tower Rearranging	9,611	-.0019001	-.0039706, .0001704	.0721	.1566	2,759	-.0048043	-.0087284, -.0008802	.1879	.8058
	Matrix pattern completion	9,670	-.0009633	-.0030134, .0010868	.3570	.3584	2,768	-.0012759	-.0051684, .0026166	.6494	.8393
	Trails A (time)	9,668	.0030946	.0010132, .005176	.0036	.0162	2,766	.0008261	-.0031038, .004756	.6301	.8393
	Trails B (time)	9,455	.0026019	.0004972, .0047067	.0154	.0554	2,742	.0023579	-.0016326, .0063484	.2171	.8058
	Trails B – A (time)	9,455	.0018798	-.0002126, .0039721	.0783	.1566	2,742	.0022594	-.0016632, .0061819	.2633	.8058
	Trails B (errors)	9,362	.001587	-.001800, .004974	.3584	.3584	2,760	.0037901	-.0025624, .0101426	.1275	.8058
	Processing Speed (comp)	9,653	.0035774	.0019994, .0051555	<.0001	.0002	2,763	.0014516	-.0014919, .0043952	.7386	.8393
	Executive Function (comp)	9,650	.0011996	-.000357, .0027562	.1309	<.2142	2,764	-.0006113	-.0035482, .0023255	.6832	.8396
	Reasoning (comp)	9,612	-.0010418	-.0027634, .0006798	.2356	.3227	2,763	-.0021697	-.0053804, .0010411	.2686	.8058
	Memory (comp)	9,964	.0009278	-.0004027, .0022584	.1717	.2576	2,848	.0007733	-.0016947, .0032413	.6198	.8393
	Global CF (comp)	14,284	.0005975	-.0004228, .0016178	.2510	.3227	3,933	.0007286	-.0012778, .0027349	.4765	.8393

Abbreviations: CF, cognitive function; CI, confidence interval; FDR, False Discovery Rate; MET, Metabolic Equivalent of Task; PA , physical activity; Uncorr, uncorrected

a: All expressed as z scores, except Prospective Memory which is expressed as an odds ratio

b: Adjusted for: alcohol binge, alcohol frequency, antihypertensive medication, apoe-e4 allele count, BMI, cardiovascular disease dx, dementia genetic risk score, diabetes dx, distance to major road, friend and family visits, gender, hdl cholesterol, head injury dx, household income, kidney disease dx, KJ of energy, ldl cholesterol, living alone status, manual work, mood disorder dx, musculoskeletal dx, neurological disorder dx, neuroticism score psychosis dx, psychotropic medication, salt added to food, smoking status, Townsend deprivation score, trauma status, waist circumference, worrier status. Also adjusted for technical covariates used with genetic risk scores.

c: Probability adjusted using the Simes-Benjamini-Hochberg method implemented in the Stata qqvalue package.

d: Processing speed composite = mean of Digit Symbol Substitution and Reaction Time (assuming non-missing on both measures)

e: Executive function composite = mean of Tower Rearranging, Trails A and Trails B completion time (assuming non-missing on two measures)

f: Reasoning composite = mean of Reasoning test and Matrix Pattern Completion (assuming non-missing on both measures)

g: Memory composite = mean of Pairs Matching, Numeric Memory and Paired Associate Learning (assuming non -missing on two measures)

h: Global CF = mean of z scores on ten tests (assuming at least two non-missing values

Appendix 3: Original research proposal

Final Approved MRP Proposal

Title: Examining the relationship between physical activity and cognitive function using a large population cohort

Student ID:

Date of submission: 27.01.2020

Version number: 011

Word count (max): 3231 (3000)

Abstract

Background: Elucidating modifiable factors that may confer protection against cognitive decline is an important area of investigation. Physical activity represents one such factor and has generated substantial evidence regarding its protective benefit. However, in epidemiological research the presence of multiple covariates has posed challenges to obtaining unconfounded estimates of associations between exposure and outcome variables, which means that we do not know true magnitude of the relationship between physical activity and cognitive function. The present study aims to address this issue by using the very large cohort dataset, UK Biobank, to construct and test a comprehensive model of this relationship.

Aims: To construct and test multivariate models in order to estimate the magnitude of the relationship between physical activity and cognitive function.

Methods: (i) A systematic review of observational studies reporting on the association between physical activity and cognitive function/decline, producing a narrative synthesis and graphical causal model.

(ii) Longitudinal analysis using the UK Biobank dataset to test the model constructed in the review phase.

Applications: Contributing to the understanding of complex relationships between demographic, clinical and lifestyle factors and cognitive health, which may have an impact on intervention and policy development.

Introduction

Cognitive function and decline

The term cognitive function (CF) describes a set of mental abilities that enable humans to acquire and use the knowledge and skills necessary to function in everyday life. A person's CF can affect many areas of their life, such as their education, employment and socio-economic status, and this wide-ranging impact has been conceptualised as the 'cognitive footprint' (Knapp, Kung, Rossor, & Stoner, 2018). Whilst some decline in CF is part of normal aging (Harada, Love, & Triebel, 2013), more pronounced decline can range from preclinical mild cognitive impairment (MCI) through to full neurodegenerative disease presentations such as Alzheimer's disease and other dementias. Such presentations result in reduced quality of life and carry significant disease burden (Gauthier et al., 2006; Nichols et al., 2019). The biggest risk factor for cognitive decline (CD) is age (Hebert, Weuve, Scherr, & Evans, 2013), and, as people are living longer, the number of people living with dementia globally has more than doubled since 1990 (Nichols et al., 2019). Whilst age itself is a risk factor not amenable to modification, there is growing evidence that a range of lifestyle factors, such as physical activity, can offer protective benefit against CD and, importantly, represent feasible targets for intervention (Baumgart et al., 2015).

Physical activity and cognitive health

Physical activity (PA) is defined as "*... any bodily movement produced by skeletal muscles that requires energy expenditure "* (Casperson, Powell and Christenson, 1985, p.1). A lack of PA is a leading risk factor for global mortality and is associated with health outcomes such as obesity, diabetes, cancer and heart disease (Lear et al., 2017). Researchers commonly conceptualise PA along a continuum, with levels ranging from 'sedentary' through to 'light', 'moderate', 'vigorous' and 'high' (Norton, Norton, & Sadgrove, 2010). Importantly, recent approaches identify sedentary behaviour (SB) as a construct which is related to PA but which can be independently associated with health outcomes (Katzmarzyk, 2010). In recent research using UK Biobank data, (Celis-Morales et al., 2019) categorised people according to their values on these two constructs to create four groups: high PA/low SB, "busy bees", (2) high PA/high SB; (3) low PA/low SB; and (4) low PA/SB time, "couch potato."

Besides these physical health outcomes, a growing literature also link PA to cognitive health (Baumgart et al., 2015) including meta-analyses of the relationship between PA and CF

which demonstrate reduced risk of CD in people with higher levels of PA (Blondell, Hammersley-Mather, & Veerman, 2014) . There are several pathways by which the benefits of PA appear to operate. At the cellular level PA appears to facilitate neurogenesis, synaptogenesis and angiogenesis (Lista & Sorrentino, 2010), which may produce changes in grey matter volume (Erickson, Leckie, & Weinstein, 2014) and white matter integrity (Smith et al., 2016). Furthermore, the cardiovascular effects of PA appear to protect against antherosclerosis (González et al., 2018).

This relationship between PA and CF is particularly relevant for people who are less physically active, such as those experiencing mental illness. Research shows that PA levels are low amongst patients with mood disorders and schizophrenia (Vancampfort et al., 2017), and even amongst those with more common conditions such as anxiety (Stubbs et al., 2017).

Cohort studies

Whilst the body of evidence linking PA to CF is already substantial, it is also recognised that, in health and epidemiological research, the presence of covariates which are associated with both exposure and outcome variables makes measuring and controlling for confounding particularly challenging (McNamee, 2003). However, the recent availability of very large cohort datasets containing a wide breadth of variables, makes constructing causal structural models which include all of the necessary covariates, whilst retaining the statistical power necessary to detect even weak effects, increasingly feasible (McIntosh et al., 2016). The present study seeks to take advantage of such opportunities in order to rigorously examine the relationship between PA and CF, whilst taking account of covariate and confounder effects.

Aims and hypotheses

The aim is to examine the relationship between PA and CF in the middle-aged-to early-old aged population of the UK Biobank cohort dataset. This will be conducted in two related phases. Phase one will consist of a systematic review of studies reporting the association between PA and CF, synthesising the relevant literature to produce an evidence-based graphical model of this relationship, including the covariates that need to be included in multivariate statistical analyses. The second phase will be informed by the outcomes of the first and will take a longitudinal approach, testing multivariate causal models of the relationship of interest. At this stage it is anticipated that the review phase will yield two hypotheses to be tested in phase two:

(i) There is a significant positive association between PA at baseline and CF at follow-up, and this will be smaller in magnitude, but still statistically significant, when adjusted for confounders.

(ii) This relationship is mediated by structural neural differences such as grey matter volume and white matter integrity.

Plan of investigation & Methods

Phase one (systematic review):

The review phase will seek to answer the following questions:

(i) In UK adults, what is the magnitude of relationship between PA and cognitive function/impairment.

(ii) What sociodemographic, clinical and other factors do studies also report as being associated with cognitive function/impairment?

Within the literature CF is commonly conceptualised using performance-based neuropsychological tests, neuroimaging measures such as structural Magnetic Resonance Imaging (MRI) and by clinical diagnoses of dementia. With feasibility in mind, the review will pertain only to studies which used neuropsychological tests to measure CF. The resulting articles will be subject to the conventional stages of screening against a set of inclusion criteria, followed by data extraction and critical appraisal.

The novel component of this review will come at the synthesis stage, at which it is proposed to construct a causal diagram, known as a directed acyclic graph (DAG). DAGs visually encode structural causal models by drawing nodes to represent variables and single-headed arrows between nodes to represent the proposed direction of causation. The DAG resulting from the review will identify and depict the variables which must be measured and controlled for to obtain an unconfounded effect size estimate of the relationship between PA and CF. To construct the relevant DAG the review process will utilise a recently developed protocol: Evidence Synthesis for Constructing Directed Acyclic Graphs (ESC-DAG) (Ferguson et al., 2019). There are three stages within the ESC-DAG protocol: 1) The findings of each included study are depicted in a DAG; 2) each DAG is evaluated using a set of causal inference criteria and adjusted if necessary; 3) the resulting set of DAGs are combined to form an

integrated DAG. The final product thus constitutes a complex multivariate model of the relationship between exposure and outcome variables.

Phase two (cohort study):

One a DAG has been constructed, this will be investigated empirically using the UK Biobank dataset.

Participants:

UK Biobank is a general population-based prospective study aiming to examine the genetic and environmental determinants of disease (Sudlow et al., 2015). The resource consists of data pertaining to 502,520 individuals who were aged 40-69 years at recruitment, (mean 56.5 years); all were registered with the NHS and living near one of 22 assessment centres.

Recruitment:

From 2006 to date, baseline and follow-up data have been collected at assessment visits lasting two hours.

Inclusion and exclusion criteria:

The subset of the overall cohort which will be used to answer each research question will be determined by the systematic review phase. This will elucidate the variables which will be necessary to include within the model, and thus only participants who have overlapping completion on the relevant measures will be included.

Measures:

The exposure variable is PA, of which there are several measures in the Biobank dataset. Previous Biobank studies have used the self-report measure the International Physical Activity Questionnaire (IPAQ) (Celis-Morales et al., 2019) whilst others have opted to use the objective measure of accelerometry data (Doherty et al., 2017). The most appropriate variables to use as primary measures in analysis will be elucidated early in the review process and will be added to the analysis protocol as soon as they are selected. The outcome variable is CF which was assessed in UK Biobank using a short, bespoke computerised battery of tests measuring memory, reasoning and reaction time. These battery of tests show moderate to good convergent validity with well-established cognitive tests and moderate to high short-term test-retest reliability (Fawns-Ritchie & Deary, 2019). Structural neural features such as

grey matter volume and white matter integrity have also been measured in Biobank (Miller et al., 2016) and will be treated as mediators in the present study. The covariates which are included in the model will be elucidated by the review process

Design:

This is a large population prospective cohort study utilising longitudinal analysis to examine the relationship between exposure and outcome variables, as well as mediating causal pathways. Of the 502,520 at baseline, around 100,000 have been followed up at least once, making longitudinal analysis possible (Conroy et al., 2019).

Research procedures:

Conducting the project will entail extracting the relevant variables from the UK Biobank database, cleaning/preprocessing the dataset and using the software programme Stata to perform analysis. Training will be provided by the project supervisor and regular collaboration with a local group of Biobank researchers will be utilised.

Data analysis

Stata will be used to conduct descriptive statistics and multiple regression models. Mediation analysis will also be conducted in Stata, to examine potential causal pathways such as structural brain differences. Alternative versions of the analyses will be conducted on a planned basis to check the sensitivity of the results to different specifications, for example using self-report vs accelerometry data to represent PA. The variables to be treated as primary and secondary will be reported in an analysis protocol prior to conducting analysis. Given the multiplicity of analyses and the very large sample size (making low p-values more likely), the threshold for statistical significance will be set at $p < 0.01$, and false discovery rate corrections will be applied if multiple models are used to test the same hypothesis. This threshold has been used in previous Biobank studies of a similar scope (Cullen et al., 2015). Given that mental health status is anticipated to be an important covariate, planned analysis of interaction effects according to mental health status will also be added to the analysis protocol once the most relevant measures have been elucidated.

Justification of sample size:

Feasibly testing such a complex model is made possible by the size of the dataset available which grants the statistical power necessary to detect even weak effects. Although

the baseline cohort is very large, there are relatively fewer participants with data on some measures (around 40,000 for neuroimaging) and the sample size with complete data on all covariates will be smaller again. Nevertheless, assuming that the smallest sample size for analysis in this study is as low as 20,000 (the neuroimaging sample, with a conservative estimate of 50% having complete data), this would still be ample to reliably estimate regression and mediation models to detect small effect sizes (Fritz & MacKinnon, 2007).

Settings:

As no data collection is required the setting of this research is non-clinical and will take place largely on the University of Glasgow campus.

Equipment:

The only equipment required is a licence for the statistical software Stata; the University has a site licence for this.

Health and safety issues:

None

Ethical issues:

All data to be used are pre-existing, and have been anonymised by UK Biobank centrally using a coding system to which researchers do not have access. UK Biobank has approval from the NHS National Research Ethics Service as a research tissue bank and separate project-specific ethical approval is not required by approved researchers using data released by UK Biobank. In accordance with the research supervisor's contract with UK Biobank, the data will be stored and analysed on a password-protected University network drive. The primary researcher will be registered as an approved researcher with Biobank before accessing data. All UK Biobank participants gave informed consent and those who withdraw are removed from the dataset on a regular basis. The findings of the current project will aim to be published in peer reviewed journals.

Ethical issues relevant to largescale cohort studies, and in particular Biobank, will be considered and discussed in the main write-up of the thesis.

Financial issues: None

Timetable:

Stage	Deadline
Proposal for blind review	January 27th 2020
Systematic review protocol complete	June 2020
Systematic review synthesis complete	February 2021
Data analysis coding development	January-May 2021
Final models ready to analyse	May 2021
Writing up	June-July 2021
Thesis submission	July 2021

Practical applications

The findings will be relevant in the context of the UK's aging population, as strategies and policies to prevent CD are developed. This research may also be relevant to those designing interventions for other clinical subgroups at risk of CD such as people with major mental illness.

References

Baumgart, M., Snyder, H. M., Carrillo, M. C., Fazio, S., Kim, H., & Johns, H. (2015). Summary of the evidence on modifiable risk factors for cognitive decline and dementia: a population-based perspective. *Alzheimer's & Dementia, 11*(6), 718-726.

Blondell, S. J., Hammersley-Mather, R., & Veerman, J. L. (2014). Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. *BMC Public Health, 14*(1), 510.

Celis-Morales, C. A., Lyall, D. M., Bailey, M. E., Petermann-Rocha, F., Anderson, J., Ward, J., . . . Sattar, N. (2019). The Combination of Physical Activity and Sedentary Behaviors Modifies the Genetic Predisposition to Obesity. *Obesity, 27*(4), 653-661. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/oby.22417

Conroy, M., Sellors, J., Effingham, M., Littlejohns, T. J., Boultwood, C., Gillions, L., . . . Allen, N. E. (2019). The advantages of UK Biobank's open-access strategy for health research. *Journal of Internal Medicine, 286*(4), 389-397.

Cullen, B., Nicholl, B., Mackay, D., Martin, D., Ul-Haq, Z., McIntosh, A., . . . Evans, J. (2015). Cognitive function and lifetime features of depression and bipolar disorder in

a large population sample: Cross-sectional study of 143,828 UK Biobank participants. *European Psychiatry*, *30*(8), 950-958.

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., . . . Owen, C. G. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. *PloS One*, *12*(2), e0169649.

Erickson, K. I., Leckie, R. L., & Weinstein, A. M. (2014). Physical activity, fitness, and gray matter volume. *Neurobiology of Aging*, *35*, S20-S28.

Fawns-Ritchie, C., & Deary, I. J. (2019). Reliability and validity of the UK Biobank cognitive tests. *MedRxiv*, 19002204.

Ferguson, K. D., McCann, M., Katikireddi, S. V., Thomson, H., Green, M. J., Smith, D. J., & Lewsey, J. D. (2019). Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *International Journal of Epidemiology*. doi:10.1093/ije/dyz220

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233-239.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., . . . Chertkow, H. (2006). Mild cognitive impairment. *The Lancet*, *367*(9518), 1262-1270.

González, H. M., Tarraf, W., Harrison, K., Windham, B. G., Tingle, J., Alonso, A., . . . Mosley, T. H. (2018). Midlife cardiovascular health and 20-year cognitive decline: Atherosclerosis Risk in Communities Study results. *Alzheimer's & Dementia*, *14*(5), 579-589.

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

Hebert, L. E., Weuve, J., Scherr, P. A., & Evans, D. A. (2013). Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*, *80*(19), 1778-1783.

Katzmarzyk, P. T. (2010). Physical activity, sedentary behavior, and health: paradigm paralysis or paradigm shift? *Diabetes*, *59*(11), 2717-2725.

Knapp, M., Kung, C. S., Rossor, M., & Stoner, C. R. (2018). Cognitive footprint: A framework to guide public policy. In *New Developments in Dementia Prevention Research* (pp. 181-193): Routledge.

Lear, S. A., Hu, W., Rangarajan, S., Gasevic, D., Leong, D., Iqbal, R., . . . Kumar, R. (2017). The effect of physical activity on mortality and cardiovascular disease in 130 000

people from 17 high-income, middle-income, and low-income countries: the PURE study. *The Lancet, 390*(10113), 2643-2654.

Lista, I., & Sorrentino, G. (2010). Biological mechanisms of physical activity in preventing cognitive decline. *Cellular and Molecular Neurobiology, 30*(4), 493-503.

McIntosh, A. M., Stewart, R., John, A., Smith, D. J., Davis, K., Sudlow, C., . . . Hassan, L. (2016). Data science for mental health: a UK perspective on a global challenge. *The Lancet Psychiatry, 3*(10), 993-998.

McNamee, R. (2003). Confounding and confounders. *Occupational and Environmental Medicine, 60*(3), 227-234.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., . . . Andersson, J. L. (2016). Multimodal Population Brain Imaging in the Uk Biobank Prospective Epidemiological Study. *Nature Neuroscience, 19*(11), 1523.

Nichols, E., Szoeke, C. E., Vollset, S. E., Abbasi, N., Abd-Allah, F., Abdela, J., . . . Asgedom, S. W. (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology, 18*(1), 88-106.

Norton, K., Norton, L., & Sadgrove, D. (2010). Position statement on physical activity and exercise intensity terminology. *Journal of Science and Medicine in Sport, 13*(5), 496-502.

Smith, J. C., Lancaster, M. A., Nielson, K. A., Woodard, J. L., Seidenberg, M., Durgerian, S., . . . Rao, S. M. (2016). Interactive effects of physical activity and APOE-ε4 on white matter tract diffusivity in healthy elders. *Neuroimage, 131*, 102-112.

Stubbs, B., Koyanagi, A., Hallgren, M., Firth, J., Richards, J., Schuch, F., . . . Lahti, J. (2017). Physical activity and anxiety: a perspective from the World Health Survey. *Journal of Affective Disorders, 208*, 545-552.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine, 12*(3), e1001779.

Vancampfort, D., Firth, J., Schuch, F. B., Rosenbaum, S., Mugisha, J., Hallgren, M., . . . De Hert, M. (2017). Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta‐analysis. *World Psychiatry, 16*(3), 308-315.

Appendices

(i) Anonymised health and safety form

APPENDIX 8.5 HEALTH & SAFETY FORM

HEALTH AND SAFETY FOR RESEARCHERS

1. Title of Project: Examining the relationship between physical activity and cognitive function using a large population cohort

2. Trainee:

3. University Supervisor: _____

4. Other Supervisor(s):

5. Local Lead Clinician: N/A

6. Participants: (age, group or subgroup, pre- or post-treatment, etc): Pre-existing cross sectional and longitudinal UK Biobank data. Participants were 502,520 UK adults aged 40-69. Around 100,000 attended follow-up assessment and their data will be used in the present study. No new data collection required for the present study.

7. Procedures to be applied (eg, questionnaire, interview, etc): No new data collection required. Participants were assessed on a variety of clinical, social and other measures in assessment sessions lasting two hours. Some participants have attended follow up sessions where more measures were administered, including brain scans.

8. Setting (where will procedures be carried out?): No data collection procedures. Analysis to be carried out on university campus.

i) General:

ii) Are home visits involved Y/N: No

8. Potential Risk Factors Identified (see chart): No

9. Potential Risk Factors Considered (for researcher + participant safety):

i) Participants:

ii) Procedures

iii) Settings

(i) Anonymised research costs and equipment form

APPENDIX 8.6 RESEARCH COSTS & EQUIPMENT

RESEARCH EQUIPMENT, CONSUMABLES AND EXPENSES

Trainee:

Year of Course: 2020

Intake Year: 2018

Please refer to latest stationary costs list (available from student support team):

Item	**Details and amount required**	**Cost or specify if to request to borrow from department**
Stationary	N/A	Subtotal:
Postage	N/A	Subtotal:
Photocopying and laser printing	N/A	Subtotal:
Equipment and software	Laptop (Dell16) borrowed from department to access STATA software	Subtotal: 0 (borrowed from department)
Measures	N/A	Subtotal:
Miscellaneous	N/A	Subtotal:
Total	**0**	**0**

For any request over £200 please provide further justification for all items that contribute to a high total cost estimate. Please also provide justification if costing for an honorarium:

(i) Plain Language Summary:

Plain English summary

ID

Date of submission: 27.01.2020

Title: Examining the relationship between physical activity and cognitive function using a large population cohort

Wordcount: 500

Background:

Cognitive function (CF) describes the range of mental abilities such as memory language, problem solving and visual ability that make it possible for humans t learn and use knowledge and skills. When CF is reduced (cognitive decline) people can find it hard to participate fully in their daily activities and their quality of life may suffer. Cognitive abilities tend to get worse with age and, as the UK has an aging population, understanding how to protect against this is an important topic for research. People who do more physical activity (PA; e.g. exercise, house work) have been found to have a lower chance of cognitive decline. This might be because PA helps keep the brain healthy. However, there are lots of other factors that are related both to PA and CF and so the true strength of the relationship between the two remains unclear. UK Biobank is a very large study of UK adults who were assessed using measures of many clinical, social and other background factors. Some of them also had brain scans. This wide range of data allows us to examine the relationship between PA and CF in a detailed way.

Aims:

By reviewing previous research, we will identify the background factors that might influence the relationship

(i) between PA and CF. We will then draw a diagram showing the relationships between these factors.

(ii) We will use statistical models to see how strong the relationship between PA and CF is, after taking account of the other important factors in the diagram. We will also test whether this relationship is explained by measures of brain health.

Methods

Participants*:* Participants were 502,520 UK adults aged 40-69. Around 100,000 attended follow-up assessment and their data will be used in the present study.

Recruitment: Participants were invited to attend assessment centres around the UK. No further recruitment is needed for the present study.

Consent: All UK Biobank participants gave informed consent and those who withdraw are removed from the dataset on a regular basis

Design of study: The study will look at whether people who reported more physical activity at their first assessment have higher cognitive function at their follow-up assessment, and whether there were differences in their brain scans.

Data collection: All participants attended a two-hour baseline assessment. Some also attended follow-up assessments, including brain scans. No new data collection is needed for the present study.

Key ethical issues: All data to be used are pre-existing and has been anonymised by UK Biobank centrally using a coding system to which researchers do not have access. UK Biobank has approval from the National Health Service (NHS) National Research Ethics Service and separate project-specific ethical approval is not required for the present study.

Practical applications and dissemination: The findings of the study may be useful in designing interventions and policy regarding physical activity and cognitive health and will be aimed to be published in peer reviewed journals.

References:

Baumgart, M., Snyder, H. M., Carrillo, M. C., Fazio, S., Kim, H., & Johns, H. (2015). Summary of the evidence on modifiable risk factors for cognitive decline and dementia: a population-based perspective. *Alzheimer's & Dementia, 11*(6), 718-726.

Conroy, M., Sellors, J., Effingham, M., Littlejohns, T. J., Boultwood, C., Gillions, L., . . . Allen, N. E. (2019). The advantages of UK Biobank's open‐access strategy for health research. *Journal of Internal Medicine, 286*(4), 389-397.