

The PREVENT Dementia programme: Baseline demographic, lifestyle, imaging and cognitive data from a midlife cohort study investigating risk factors for dementia.

Craig W. Ritchie^{1,2}, Katie Wells¹, Sarah Gregory¹, Isabelle Carriere³, Samuel O. Danso¹, David Driscoll⁴, Maria-Eleni Dounavi⁵, Robert Hillary¹, Ivan Koychev⁶, Brian Lawlor⁷, Su Li^{5,8}, Audrey Low⁵, Elijah Mak⁵, Paresh Malhotra^{9,10,11}, Jean Manson^{1,12}, Riccardo Marioni¹³, Lee Murphy¹⁴, Lorina Naci^{7,15}, John T O'Brien⁴, William Stewart^{16,17}, Graciela Muniz-Terrera^{1,18} and Karen Ritchie^{1,3} on behalf of the PREVENT Dementia programme group

¹*Edinburgh Dementia Prevention, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK*

²*Scottish Brain Sciences, Edinburgh, UK*

³*INM, Université de Montpellier, INSERM, 34000 Montpellier, France*

⁴*PREVENT Participants Panel Member, UK*

⁵*Department of Psychiatry, School of Clinical Medicine, University of Cambridge, Cambridge, UK*

⁶*Department of Psychiatry, University of Oxford, Oxford, UK*

⁷*Global Brain Health Institute, Trinity College Dublin, Ireland*

⁸*Department of Neuroscience, University of Sheffield, Sheffield, UK*

⁹*Imperial College London and the University of Surrey, UK Dementia Research Institute Care Research and Technology Centre, London, UK*

¹⁰*Brain Sciences, Imperial College London, London, UK*

¹¹*Imperial College Healthcare NHS Trust, Clinical Neurosciences, Charing Cross Hospital, London, UK*

¹²*Roslin Institute, University of Edinburgh, Edinburgh, UK*

¹³*Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK*

¹⁴*Edinburgh Clinical Research Facility, University of Edinburgh, Edinburgh, UK*

¹⁵*Trinity College Institute of Neuroscience, School of Psychology, Trinity College Dublin, Dublin, Ireland*

¹⁶*Department of Neuropathology, Queen Elizabeth University Hospital, Glasgow, UK*

¹⁷*School of Psychology and Neuroscience, University of Glasgow, Glasgow, UK*

¹⁸*Ohio University Heritage College of Osteopathic Medicine, Ohio University, Ohio, USA*

Corresponding author: Craig Ritchie (Craig.Ritchie@ed.ac.uk)

Professor Craig Ritchie

Edinburgh Dementia Prevention,

Outpatient Department Level 2,

University of Edinburgh, Edinburgh

EH4 2XU

Abstract (365 words, limit 400)

PREVENT is a multi-centre prospective cohort study in the UK and Ireland that aims to examine mid-life risk factors for dementia, identify and describe the earliest indices of disease development. The PREVENT dementia programme is one of the original epidemiological initiatives targeting midlife as a critical window for intervention in neurodegenerative conditions. This paper provides an overview of the study protocol and presents the first summary results from the initial baseline data to describe the cohort.

Participants in the PREVENT cohort provide demographic data, biological samples (blood, saliva, urine and optional cerebrospinal fluid), lifestyle and psychological questionnaires, undergo a comprehensive cognitive test battery, and are imaged using multi-modal 3T magnetic resonance imaging (MRI) scanning, with both structural and functional sequences. The PREVENT cohort governance structure is described, which includes a steering committee, a scientific advisory board and core patient and public involvement groups. A number of sub-studies which supplement the main PREVENT cohort are also described.

The PREVENT cohort baseline data includes 700 participants recruited between 2014 and 2020 across five sites in the UK and Ireland (Cambridge, Dublin, Edinburgh, London and Oxford). At baseline, participants had a mean age of 51.2 years (range 40-59, SD ± 5.47), with the majority female ($n=433$, 61.9%). There was a near equal distribution of participants with and without a parental history of dementia (51.4% vs 48.6%) and a relatively high prevalence of *APOE* $\epsilon 4$ carriers ($n=264$, 38.0%). Participants were highly educated (16.7 \pm 3.44 years of education), were mainly of European Ancestry ($n=672$, 95.9%) and were cognitively healthy as measured by the Addenbrookes Cognitive Examination-III (ACE-III) (Total score 95.6 \pm 4.06). Mean white matter hyperintensity (WMH) volume at recruitment was 2.26 \pm 2.77 ml (median = 1.39ml), with hippocampal volume 8.15 \pm 0.79ml. There was good representation of known dementia risk factors in the cohort.

The PREVENT cohort offers a novel dataset to explore midlife risk factors and early signs of neurodegenerative disease. Data are available open access at no cost via the Alzheimer's Disease Data Initiative (ADDI) platform and Dementia Platforms UK (DPUK) platform pending approval of the data access request from the PREVENT steering group committee.

Introduction

The PREVENT dementia programme (PREVENT) was initiated in 2014 as a single-site study based in West London. It has subsequently expanded to become a multi-centre study, opening sites in Edinburgh (2015), Oxford (2017), Cambridge (2017) and Dublin (2018). The aims of PREVENT are to profile midlife risk factors for later-life neurodegeneration and to identify the earliest indices heralding neurodegenerative disease in advance of clinically diagnosable dementia (particularly Alzheimer's disease (AD)). The original baseline protocol for the pilot site is described elsewhere [1, 2], with this current paper serving to provide an update on the protocol, detailing a multitude of sub-studies supplementing the main study, and provide an overview of the baseline dataset.

Recruitment

Participants were recruited via various methods. Initially participants were recruited as family members of patients at NHS memory clinics at the participating sites and through local dementia research registers. Following this, family and friends of participants were invited to participate and recruitment took place via word of mouth. The Join Dementia Research (JDR) platform (www.joindementiaresearch.nihr.ac.uk) was also utilised to recruit participants along with some participants registering their interest to participate through the PREVENT dementia website (www.preventdementia.co.uk).

Across the five centres, 700 participants have completed baseline assessments and first follow-up (visit 2) around two years after baseline. A second wave of follow up visits is underway at the London site and planned at the other centres, re-assessing participants at five to eight years post-baseline.

PREVENT has also collaborated with a number of sister projects since its inception. The TriBEKa collaboration (<http://tribeka.org/>) was established in 2017 between the Barcelona Beta Brain Research centre (the ALFA project [3]), University of Edinburgh (PREVENT) and the Karolinska Institute, with the aim of supporting ongoing cohorts of healthy adults at a spectrum of risk for dementia with a focus on neuroimaging data collection. The aim of the collaboration is to harmonise neuroimaging datasets where appropriate and support the addition of rich neuroimaging data from the cohorts to the Global Alzheimer's Association

Interactive Network (GAAIN) and Alzheimer's Disease Data Initiative (ADDI) portals for worldwide academic access. In addition to TriBEKa, PREVENT was associated with the European Prevention of Alzheimer's Dementia (EPAD) programme [4-6]. The EPAD Longitudinal Cohort Study included a wide spectrum of participants at differing levels of risk for AD. In addition to being a recruitment source as a parent cohort, PREVENT influenced the design of the EPAD LCS protocol. Importantly the participant involvement experience from PREVENT ensured this became a core pillar of EPAD, with significant impact on the study success reported [7]. Focus groups involving PREVENT participants also explored ethical aspects of the EPAD project before initiation which was developed into a work package focusing on ethics within the EPAD project [8, 9].

Materials and Methods: The PREVENT Dementia Protocol

Ethics

Multi-site ethical approval was granted by the UK London-Camberwell St Giles National Health Service (NHS) Research Ethics Committee (REC reference: 12/LO/1023, IRAS project ID: 88938), which operates according to the Helsinki Declaration of 1975 (and as revised in 1983). A separate ethical application for Ireland was submitted for the Dublin site, was reviewed and given a favourable opinion by Trinity College Dublin School of Psychology Research Ethics Committee (SPREC022021-010) and the St James Hospital/Tallaght University Hospital Joint Research Ethics Committee. All substantial protocol amendments have been reviewed by the same ethics committees and favourable opinion was granted before implementation at sites. All sub-studies referred to have individual ethical applications and favourable opinions.

Demographics

Participants self-reported demographic information via interview with a researcher during each study visit. Demographic data was gathered to provide descriptive data on the cohort and to include a number of known risk and confounding factors for neurodegeneration. The demographic data include date of birth, sex, years of education, family history of dementia (including subtype, age of onset and age of death where known), occupation, postcode and handedness.

Biosamples

All participants were asked to provide blood, urine and saliva samples, with an option to undergo a lumbar puncture for cerebrospinal fluid. Approximately 50 ml of blood were collected from overnight fasted participants. Clinical samples were analysed immediately for standard biochemistry and haematology measures at local laboratories, with results entered into the participant database. Research samples were processed and prepared for long term storage as plasma, buffy coat, serum and whole blood samples (for DNA extraction) and stored at -80°C.

Saliva samples were also collected from all participants on two different days across eight time points. The first day of sample collection, termed a controlled stress day, was the day

of their study visit when the clinical and cognitive assessments were completed. Participants were asked to complete the second day of samples (requested to be within a week of the first day of sampling but up to one month from the first sample day) on a quieter day at home (quieter day recommended to be a day spent mainly at home where participants did not envisage encountering any significant stressors). Stimulated saliva was collected using Salivette® tubes (Sarstedt, Germany) Cortisol collection tubes with a synthetic swab. Samples were returned to the research unit after completion and stored at -20°C.

A 12-hour overnight urine collection was also completed by all participants which was then processed and prepared for long-term storage. 40 ml of urine was extracted and stored for each participant, 20 ml as standard and 20 ml acidified with hydrochloric acid, then stored at -80°C.

All processed biosamples are stored at the Scottish Brain Health Bioresource, The Roslin Institute, University of Edinburgh.

Genetic data

Genomic DNA from PREVENT participants was isolated from whole blood samples using a Nucleon Kit (GenProbe) with the BACC3 protocol. DNA samples were re-suspended in 1 ml TE buffer pH 7.5 (10mM Tris-Cl pH 7.5, 1mM EDTA pH 8.0). The yield of the DNA was measured using picogreen. *APOE* genotyping was performed using Taqman polymerase chain reaction (PCR) genotyping and the QuantStudio 12K Flex system ($n=696$). The final volume was 5 µl using 20 ng of genomic DNA, 2.5 µl of TaqMan Master Mix, and 0.125 µl of 40× Assay by Design or 0.25 µl of 20× Assay on Demand Genotyping Assay. The cycling parameters were 95° for 10 minutes, 40 denaturation cycles at 92° for 15 seconds and annealing/extension at 60° for one minute.

Six hundred and ninety-six samples underwent genome-wide genotyping on the Infinium™ Global Screening Array-24 v3.0 BeadChip ($n = 730,059$ loci) and scanned on an Illumina iScan platform. Genotypes were called automatically using GenomeStudio Analysis software v2011.1 and Quality control (QC) was performed using PLINK v1.9 [10]. Samples and probes were removed based on the following criteria: genotype call rate (<95%), SNP missingness

(>1%, --geno 0.01), sample missingness (>1%, --mind 0.01), Hardy-Weinberg equilibrium (P-value < 1×10^{-6} , --hwe 1e-6), minor allele frequency (<0.5%, --MAF 0.005), and heterozygosity outlying values (F statistic > 3 SDs). In total, 647 samples and 515,602 SNPs passed QC. We further identified and removed 31 individuals related to another cohort member. To protect against sex imbalance in the sample, the first exclusion criterion was to remove females from male-female pairs. The second criterion was to exclude the individual with the poorer genotype call rate in male-male or female-female pairings. We also removed 20 ancestry outliers (i.e. of non-European ancestry), leaving 596 samples in our most stringent dataset. Relatedness was estimated via an identity-by-descent coefficient ≥ 0.1875 , which represents the halfway point between second and third degree relatives. Ancestry outliers were identified by principal component analyses on the PREVENT genotype dataset merged with HapMap III reference data. PREVENT genotypes were also imputed against European sample data from the Haplotype Reference Consortium (HRC) build release 1.1 (GRCh37/hg19), 1000 Genomes Phase 3 (version 5), and the TOPMed r2 reference panel [11-13]. There were 8,651,773, 9,803,244 and 10,082,029 imputed, autosomal SNPs for HRC, 1000G and TOPMed panels, respectively (imputation quality score $R^2 \geq 0.6$ and $MAF \geq 0.005$).

Physical examination

As part of the clinical assessment participants underwent a physical and neurological examination, an electrocardiograph (ECG), spirometry (removed during the Covid-19 pandemic), vital signs and anthropometric measurements (height, weight, leg length, waist, hip, and neck measurements).

Imaging

Six hundred and sixty six brain imaging datasets were collected using 3T Siemens Magnetic Resonance Imaging (MRI) scanners (specific models: Verio, PRISMA, Prisma Fit, Skyra). Image processing of the T1-weighted structural scans was carried out using FreeSurfer version 7.1.0 following correction for field inhomogeneities using the N4 algorithm [14, 15]. In particular, using the *recon-all* pipeline, global volumetrics, cortical thickness and hippocampal volume were measured. Manual corrections were conservatively applied to the *recon-all* outputs where appropriate by trained operators. Structural MRI scans were also used for the quantification of cerebral small vessel disease (SVD) markers. White matter

hyperintensity (WMH) volumes were quantified from lesion masks obtained from FLAIR MRI using an automated script on SPM8. Lesion maps obtained from the segmentation procedure were used as starting points for manual WMH delineation. Details on the procedures involved on all volumetric analyses have been described previously [16-19]. The diffusion weighted imaging (DWI) datasets were first carefully examined for sufficient coverage and minimal eddy-current distortions and preprocessed using MRTRIX (<https://www.mrtrix.org/>) and FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT/UserGuide>). Diffusion tensor imaging (DTI) parameters such as fractional anisotropy (FA) and mean diffusivity (MD) are derived using the *dtifit* function in FSL. Please refer to Table 1 and Figure 1 for acquisition parameters of all scan sequences and details on SVD quantification methods, respectively.

Resting state BOLD functional MRI (fMRI) of approximately 10 minutes was acquired from each participant who was instructed to keep their eyes closed and not to think about anything specific. Participants also completed a task based (fMRI), which was divided into two parts separated by approximately 25-30 minutes. All participants had normal or corrected normal vision and were provided with verbal instructions and an opportunity to practice responding before engaging in the task.

Part 1 (approximately six minutes duration)

Participants were shown 37 indoor and 38 outdoor images (75 in total) randomly selected from a total of 50 indoor and 50 outdoor images. The images stayed on the centre of the screen for three seconds. The participant then had up to two seconds to respond by pressing one of the two buttons to indicate whether the image they saw was an indoor or outdoor scene. Participants were not informed they would be tested for their memory of these images at this stage.

Part 2 (approximately eight minutes duration)

After a delay of approximately 25-30 minutes, participants were presented with 100 images (50 indoor scenes and 50 outdoor scenes) in pseudorandom order. 75 of these images were already presented in the first part of the task, while 25 were new images. Each image was

presented again for three seconds and participants had up to two seconds to indicate whether this is a previously seen or new image.

Analysis of data generated by the fMRI task was conducted using SPM, RSA toolbox [20] and in-house MATLAB scripts.

	TR (ms)	TE (ms)	Flip angle	Voxel size (mm ³)	Slices	Additional comments
T1-weighted	2300	2.98	9°	1 x 1 x 1	160	MPRAGE
T2-weighted	1500	80	150°	0.69 x 0.69 x 4.0	32	----
FLAIR	9000	94	150°	0.43 x 0.43 x 4.0	27	----
SWI	28	20	15°	0.72 x 0.72 x 1.2	72	----
ASL	2500	11	90°	3.0 x 3.0 x 3.0	14	PICORE; 50 averages Two variants differing in bolus duration (700 and 1675ms)
DTI	11700	90	90°	2.0 x 2.0 x 2.0	63	b= 0; 1000 s/mm ²
MRS	2000	30, 33 and 40	90°	20 x 20 x 20	Single voxel	PRESS with and without water suppression (96 and 16 averages respectively)
BOLD fMRI	2000	30	80°	3.0 x 3.0 x 3.0	35	Resting state: 330 repetitions Task1: 188 Task2: 250
Coronal hippocampal T2-weighted	6420	11	160°	0.41 x 0.41 x 2.0	20	----

Table 1. MRI acquisition parameters in the PREVENT-Dementia programme.

Abbreviations: ASL = arterial spin labelling, BOLD = blood oxygen level dependent, DTI = diffusion tensor imaging, fMRI = functional magnetic resonance imaging, MRS = magnetic resonance spectroscopy, TR = repetition time, TE = echo time, FLAIR = fluid-attenuated inversion recovery, SWI = susceptibility-weighted imaging.

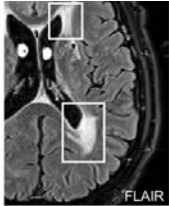
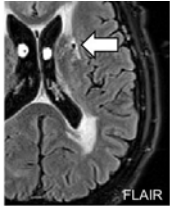
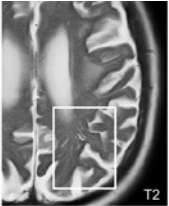
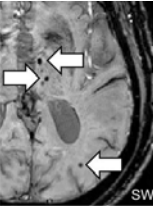
				
	White matter hyperintensities	Lacunes	Enlarged perivascular spaces	Cerebral microbleeds
What are they?	Patchy or diffuse lesions thought to represent axonal loss and demyelination	Focal subcortical infarcts caused by occlusion of perforating arteries	Microscopic fluid-filled spaces surrounding perforating vessels of the brain that become visible when dilated. Also referred to as Virchow-Robin spaces	Small foci of chronic accumulation of blood products in brain tissue. Also referred to as microhaemorrhages
MRI sequence and appearance (+) hyperintense/bright (-) hypointense/dark	FLAIR (+)	T1-weighted (-) T2-weighted (+) FLAIR (-)	T2-weighted (+)	SWI (-)
Typical size	Variable	3-15 mm	<3 mm	2-5 mm Up to 10 mm
Shape	Irregular Punctate/Confluent	Round / Ovoid	[Axial view] In centrum semiovale: rounded / linear In basal ganglia: round/ ovoid, cyst-like	Round / Ovoid
Method of quantification	Semi-automated quantification of volumes + Fazekas rating	Manual identification with cross-verification in T1, T2 and FLAIR scans	EPVS rating scale (range from 0-4)	Manual identification according to Microbleed Anatomical Rating Scale (MARS)

Figure 1: Imaging markers of cerebral small vessel disease taken from Low et al. with permission (2022) [21]

Cognitive assessments

Participants completed a battery of cognitive assessments, in particular, focusing on cortical and sub-cortical brain regions hypothesised to be first affected in neurodegenerative disease, with a preference for early stages of AD. All experimental cognitive measures were selected by experts in the neuropsychology of ageing due to their ability to detect very subtle quantitative and qualitative changes in cognition.

COGNITO: A computerised battery of tasks, designed to detect the widest possible range of cortical and sub-cortical deficits. The battery taking approximately 45 minutes to complete includes the following sub-tests: reaction time; phonemic and syntactic comprehension; auditory and visual attention; visuospatial associative learning, working memory; immediate, delayed and cued visual and verbal recall; conceptual sequencing; naming; semantic access; vocabulary [22]. A tactile screen is used to capture response latencies and

qualitative aspects of performance such as perseveration, proactive interference and visual field neglect.

The Four Mountains Test (FMT): Administered by a tablet device the FMT assesses linkage between episodic and spatial functions of the hippocampus permitting representation of spatial information in an allocentric form and hence encoding of the context in which events occur [23]. Computer-generated landscapes comprised of four hills (of varying shape and size) surrounded by a distant semi-circular mountain range are presented with a sample image for 10 seconds following which the subject is immediately presented with four alternative images. One of which (the target image) shows the same topography as the sample image, seen from a novel viewpoint, from which they must identify the target image by pressing a key. Non-spatial features (lighting, vegetation, weather conditions) of both target and foil landscapes are varied between presentation and testing, such that transient local features of the image cannot be relied on to solve the task. The task takes approximately 15 minutes to complete

The National Adult Reading Test (NART): A 50-item word pronunciation test providing an indicator of premorbid intellectual functioning taking 10 minutes to complete [24].

The Virtual Supermarket Trolley (VST): The VST is sensitive to deterioration in the precuneus, retrosplenial cortex and entorhinal connections and measures egocentric spatial orientation (as opposed to allocentric) through presentation of 14 video vignettes in an ecological virtual supermarket from a first person perspective [25]. A route is taken through a supermarket in which the participant is behind the trolley and involves series of 90° turns and at the end the subject is required to point in the direction of the entry. The task is also administered through a computerized tablet device but responses are recorded on paper by a researcher.

The Visual Short-Term Memory Binding Test (VSTMBT); Assesses memory binding abilities using combinations of shapes and colours on a computerised assessment taking approximately 15 minutes to complete. The test has been shown to predict familial AD 10-

15 years prior to the onset of clinical symptoms and is therefore a critical test to be used in this group [26].

ACE-III: The Addenbrookes Cognitive Examination III (ACE-III) provides a brief screen of possible memory, attention, fluency, language and visuospatial disabilities. The test was included following the pilot data collection to include a clinically validated measure of cognition to ensure there were no pre-existing signs of cognitive impairment which would exclude participants from the study [27]. The test is a pen and paper assessment, taking approximately 15 minutes to complete.

Self-report questionnaires

Participants completed a series of self-report questionnaires covering multiple lifestyle and risk factor domains. These included questionnaires on pregnancy and menstruation, the Lifetime of Experiences Questionnaire [28], history of educational attainment, physical activity [29], musical expertise, depression (Center for Epidemiologic Studies- Depression Scale (CES-D)) [30], anxiety (State-Trait Anxiety Inventory (STAI)) [31], sleep (Pittsburgh Sleep Quality Index (PSQI)) [32, 33], resilience (Connor-Davidson Resilience scale) [34], stressful life events (Life Stressor Checklist-Revised (LSC-R)) [35], traumatic brain injury (Brain Injury Screening Questionnaire (BISQ)) [36] and diet (Scottish Collaborative Group Food Frequency Questionnaire (SCQ-FFQ)) [37].

Sub-studies

Alongside the main PREVENT study, various researchers from institutions across the UK, Ireland and France have joined as collaborators to recruit PREVENT participants to additional sub-studies (Table 2). Data from these studies will be added to the main PREVENT database following embargo periods.

Retinal Imaging in PREVENT

Participants at the Edinburgh site are invited to undergo a retinal imaging protocol. Imaging the retina is a non-invasive and relatively easy process, making it an ideal area to investigate for translation to clinical practice. Evidence is accumulating that implicates microvasculature in neurodegenerative disease aetiology [38-41], with drusen on the retina more prevalent in

AD [42]. The retinal imaging sub-study aims to investigate retinal imaging measures in relation to dementia risk in PREVENT.

Amyloid Imaging in PREVENT (AIP) study

Up to 200 PREVENT participants are being invited to take part in the Amyloid Imaging in Prevent (AIP) study. This study involves undergoing a PET-CT scan to measure amyloid deposition in the brain.

Tau Imaging

A sub-group of 100 PREVENT participants in the Amyloid Imaging in PREVENT study are also invited to join a tau imaging study, to additionally measure levels of tau in their brain using PET.

The Entorhinal CoRtex Structure and Function in PREVENT (ENCRYPT) Study

The aim of the ENCRYPT study is to investigate whether the structure and function of the entorhinal cortex may be impaired in mid-life in those who may be at a higher risk of future AD. One hundred participants completed virtual reality tasks, with a subset of 55 participants additionally completed a 7T MRI brain scan.

Football and Rugby cohort

In addition to the main cohort the PREVENT programme is being further developed through recruitment of participants who are ex-professional football (Brain Health Outcomes in former Professional and Elite athletes; BrainHOPE) or rugby (PREVENT-Rugby Footballer Cohort; PREVENT-RFC) players. In total 210 ex-professional or elite players (male and female) will be recruited allowing for focused analyses exploring specific early indicators of disease for players from these sports and comparing to non-sports players from the wider PREVENT cohort.

Study name	Lead Researcher	Study description
Linguistic markers of future risk for Alzheimer's disease	Prof Alison Wray, University of Cardiff	Online assessment where language use was analysed to investigate whether linguistic markers could identify potential risk for future AD. 179 participants completed the baseline assessment, and 35 were followed-up two years later to identify any changes in language use across this time period.
Approaches to the Communication of Alzheimer's disease risk (ACAR study)	Dr Richard Milne, University of Cambridge	Focus groups held with research participants investigating attitudes to communication of future risk of dementia. 16 PREVENT participants were recruited alongside additional volunteers recruited from other research studies across Europe. The focus groups were structured to explore participants' interest in learning about their AD risk and what their preferences were around disclosure. [43].
PREVENT-Elicitation of Dialogues (PREVENT-ED) study	Dr Sofia De La Fuente Garcia	Collected speech data from 43 participants enrolled at the Edinburgh site while they engaged in a cognitively stimulating task. Audio recordings were processed for speech features and machine learning methods used to test for associations between these features and risk factors collected in the PREVENT cohort [44].
Mobile-technologies for the Assessment of Cognition (MTAC) study	Dr Ivan Koychev, University of Oxford	35 participants recruited from PREVENT Oxford site. Explored usability of a smart-phone based application to track cognition and function and positional technology using interactions between a smartwatch and Bluetooth beacons positioned around the homes to assess the level of function, activity and ability to navigate the environment.
<i>Neureka</i>	Dr Claire Gillan, Trinity College Dublin	94 PREVENT Participants completed cognitive assessments via a mobile phone based application to assess the validity of these tests in comparison to gold standard in person assessments.
Oral Health in PREVENT	Prof Angus Walls, University of Edinburgh	Pilot study conducted at the Edinburgh and Dublin PREVENT sites investigating periodontal disease and future risk for dementia. Participants invited for a dental examination including a dental x-ray and provided plaque and saliva samples.
Sunrise in PREVENT study	Prof Yves Dauvilliers, University of Montpellier	Participants were invited to wear a sleep activity recording device attached to their chin which measures mandibular movements to monitor sleep behaviour and aid the diagnosis of obstructive sleep apnoea.
Barriers, Facilitators and motivators to dementia prevention research	Dr Laura Booi, Leeds Beckett University	Aimed to explore what facilitates participation in dementia prevention research and what might be barriers. Recruitment was targeted to those from seldom heard groups in research to try and understand what challenges there might be to research participation. Interviews were conducted with 19 participants and analysed using thematic analysis.
Fear about memory loss in mid-life	Dr Francesca Farina, Northwestern University, USA	Participants completed an online assessment exploring fear of memory loss in midlife. Various scales related to fear and avoidance were conducted by participants to understand the level of fear certain individuals may have about dementia whilst in midlife and whether this may impact their social behaviour.
Auditory Attention in Cognitive disorder (AudCog)	Dr Meher Lad, Newcastle University	Participants invited to complete an online auditory assessment using multiple tasks to assess auditory function (pure-tone audiometry, speech-in noise perception task, auditory figure ground task, auditory working memory task, auditory sequence learning task).

Table 2. Overview of sub-studies which have recruited PREVENT participants

Operational organisation

Steering committee

The PREVENT programme is managed by a Steering group committee comprising of the Principal investigators from each site, two representatives from the PREVENT Participants' panel (detailed below), the study statistician and other academics from relevant disciplines with a key role in managing the research programme. Meetings are held quarterly to discuss study progress, funding plans, study developments, such as any new sub-studies in the pipeline, as well as any other core study business. This steering group committee also reviews and approves data and sample access requests and project proposals.

Participant and Public Involvement (PPI) panels

Participant and public involvement (PPI) has been at the core of the PREVENT Dementia programme since its inception, with the establishment of a participant panel during the study design phase. The original participant panel was set up to support the London centre of the project through the pilot phase, and two members of the panel were elected to sit on the steering committee. As the project has expanded to multiple centres and countries, the original panel have moved to support the wider project. The participant panel set up is well-described elsewhere [45]. Briefly, the core panel consists of seven participants and one non-participant who meet with the Chief Investigator (CWR) and National Coordinator (KW) quarterly. The aim of the panel meetings is to discuss project progress, future aims, sub-studies and proposed analyses. To date the panel has had a significant and positive impact on the project, supporting with recruitment, inclusion of additional sub-studies, understanding the participant experience and contributing to the future of the study [45]. In addition to this a participant panel has been established at the Edinburgh site, to support with the large number of sub-studies active at that centre. The Edinburgh panel was established in 2019 via advertisements to all active participants and have met once in person and multiple times online. The panel have supported reviews of sub-studies and supported staff to make decisions about approaches for recruitment to aforementioned sub-studies. The panels also help to co-develop any events aimed at participants such as annual conferences to share study findings.

Data Management and Quality control (QC)

As a first QC step study monitoring is carried out on a regular basis by the National Coordinator as delegated by the sponsor and PI, study documentation is reviewed for errors and omissions at all study sites. The data is entered electronically onto the REDCap data management system [46]. REDCap is a web-based software platform designed to support research data capture and management, hosted at the University of Edinburgh and managed by the study team. The system generates queries for research staff at the point of data entry. The creation of the project into REDCap replicates the same structure as the Case Report Form (CRF). This ensures that all the information from the CRF is captured and stored properly when it is entered electronically. The design of the project into REDCap includes several countermeasures to ensure the best possible quality of the extracted data. Within REDCap, fields that contain important values are designed to be mandatory to store the necessary data, with flags alerting users to any omissions. Field restrictions have been implemented to avoid mistakes and prevent data entries from being inaccurate. For example, dates are checked for values that are outside of specific ranges, as well as being in the expected form, such as integer, date, time, text. Branching logic ensures certain fields remain hidden to research staff if the participant was not eligible to answer specific questions, avoiding the possibility of entering data in inaccurate fields.

Raw imaging data are transferred and backed-up in the University of Cambridge XNAT platform. The unprocessed MRI data along with derived imaging maps and quantified neuroimaging measures are reviewed on a case-by-case basis by the PREVENT-Dementia imaging team in Cambridge. Visual assessments along with derived quality measures capturing signal and contrast to noise ratio are employed to assess image quality. Where appropriate (e.g. imaging artifacts) scans or derived measures are excluded from further analysis. All scans were reviewed at each site and any incidental findings were reported back to the study team, who then fed back to the participants, and where relevant, their primary care practitioners.

The raw MRI DICOM data were anonymized to remove identifiable information contained in imaging file headers using the DICOM confidential software [47]. Data integration and pseudoanonymisation protocols consistent with data protection principles outlined in the

UK General Data Protection Regulation (GDPR) are finally applied prior to data release for research.

Data access

Open data access is an underpinning principle of the PREVENT dementia programme, with ambitions that data collected through this study will be critical to understanding brain health in the mid-life period. The dataset has already been highly requested and resulted in several publications from outside the core study team (see Table 3). The addition of the data to the ADDI platform is anticipated to increase the accessibility and use of this novel dataset especially to Low and Middle-Income countries (LMICs).

Data requests and publications	Number
Data access requests	
Internal to consortium	62
External to consortium	44
Sample access requests	
Internal to consortium	3
External to consortium	3
Publications relating to PREVENT as of 1 st July 2023	33 (18 imaging, 7 cognition, 8 other topics) See https://preventdementia.co.uk/publications/

Table 3. Data and sample access requests from study inception to July 2023 as well as publications arising from the PREVENT cohort from study inception to July 2023.

Results: Description of PREVENT v700.0 dataset

Demographics and APOE ϵ 4 descriptive statistics

The baseline dataset includes 700 participants, with the majority of participants recruited at the Edinburgh (n=222, 31.7%) and London sites (n=210, 30.0%) (Table 4).

Site	N (%)
Cambridge	100 (14.3)
Dublin	100 (14.3)
Edinburgh	222 (31.7)
Oxford	68 (9.7)
London	210 (30.0)

Table 4. Table of number of participants in final dataset from each site.

There is a predominance of female participants (n=433, 61.9%) at all sites except Dublin (Supplementary Figure S1), with a nearly even split on those with and without parental history of dementia (has parental history, n=360, 51.4%) resulting from the targeted recruitment method used. Participants had a mean age of 51.17 years (± 5.47) at baseline, were highly educated (mean: 16.69 \pm 3.44 years) and had high prevalence of *APOE* ϵ 4 carriers (n=264/694 [38.0%] of which 34 [4.9%] homozygotes). There were no differences in *APOE* ϵ 4 by site (Supplementary Figure S1). The cohort mainly included participants of European Ancestry (n=672, 95.96%). Participants are categorised into high (positive parental history and *APOE* ϵ 4 carrier), medium (either positive family history or *APOE* ϵ 4 carrier) and low (neither family history nor *APOE* ϵ 4 carrier) risk groups, with an approximately even split across the three risk groups (high: 232, 33.4%; medium: 305, 43.9%; low: 157, 22.6%). Full descriptive details are available in Table 5.

Cognitive domains overview

Cognitive impairment was screened for by the ACE-III (note results not available for n=233 participants at Baseline as incorporated via a protocol amendment after these visits were complete). Mean cognitive scores for the cohort and by risk group are presented for the ACE-III, COGNITO tasks, FMT and VST in Table 5, and by sex in Table 6. Linear regression models were used to explore significant associations between cognitive scores and either risk group or sex.

Variable	Mean (SD) / N (%)	
Education (years)	16.69 (± 3.44) Range: 0-38 years	
Parental History of dementia	360 (51.4%)	
Sex (Female)	433 (61.9%)	
Age (years)	51.17 (± 5.47) Range: 40-60*	
	No family history	With family history
<i>APOE</i> ϵ 4 non-carrier	Low-risk N = 232, 33.4%	Medium-risk N = 198, 28.5%
<i>APOE</i> ϵ 4 carrier	Medium-risk N = 107, 15.5%	High-risk N = 157, 22.6%

Table 5. Demographics of total cohort

*note two participants were aged 60 at the time of Baseline demographic data collection, excluded from analysis. Two participants also excluded from analyses with missing data on total years of education.

COGNITO

There were no statistically significant differences on any of the COGNITO tasks by risk group in models fully adjusted for sex, age, years of education, parental history of dementia and *APOE* ϵ 4 (see Table 6). There were significant differences in performance between male and female participants on three of the COGNITO tasks. Male participants performed better than female participants on a visual and auditory attention task (male: 9.90 (± 0.33); female: 9.78 (± 0.51), β : 0.13, $p < 0.001$). Female participants performed better on a memory task of face and name recognition (female: 5.69 (± 2.02); male: 4.57 (± 2.08), β : -1.05, $p < 0.001$) and a language task of semantic verbal fluency compared to male participants (female: 17.44 (± 3.97); male: 14.68 (± 3.88), β : -2.63, $p < 0.001$). Higher scores on the name-face recognition task (memory domain) were associated with younger age (β : -0.06, SE: 0.01, $p < 0.001$). More years of education were associated with four tasks (visual and auditory attention (attention domain): β : 0.01, SE: 0.005, p : 0.02; name-face recognition (memory domain): β : 0.07, SE: 0.02, p : 0.002; semantic verbal fluency (language domain): β : 0.27, SE: 0.04, $p < 0.001$; phonemic verbal fluency (language domain): β : 0.14, SE: 0.04, p : 0.002). A full breakdown of

the percentiles for each sub-test in COGNITO by full cohort, by sex and education, and by education and dementia risk score is provided in Appendix One, Supplementary Tables 1 and 2.

FMT and VST

There were no differences in performance on either the FMT (n=459) or VST (n=453) by risk group or age. Higher scores on the FMT were associated with a higher number of years of education (β : 0.11, SE: 0.03, $p < 0.001$). Male participants performed better on the VST compared to female participants (male: 10.97 (± 1.79); female 10.14 (± 2.38), β : 0.90, $p < 0.001$).

ACE-III

Data from the ACE-III are available for 464 participants at the baseline visit, this assessment has added part way through the baseline data collection which is why this data is not available for all participants at baseline. There were no statistically significant differences between the groups in linear regression models fully adjusted to include sex, age, years of education, parental history of dementia and *APOE* ϵ 4. Female participants performed better than male participants on the ACE-III, although both mean scores were within the normal range (female: 96.17 (± 3.87); male: 94.73 (± 4.17), $p < 0.001$). When applying a clinical cut-off of 88 (recommended dementia caseness cut off for sensitivity [48]) there are 30 participants; seven in the low-risk group (age range 40-59, 42.9% female); twenty in the medium-risk group (age range 41-59, 40% female) and three in the high-risk group (age range 49-55, all male). When using a clinical cut-off of 82 (recommended dementia caseness cut-off for specificity [48]), two participants score at or below this (one in the low risk and one in the medium-risk group).

Domain	Cognitive test	Total (mean (SD)) Range	Low-risk (mean (SD))	Medium-risk (mean (SD))	High-risk (mean (SD))	Unadjusted model β , SE, p (Comparator: High Risk Group)	Fully Adjusted model β , SE, p (Comparator: High Risk Group)
General cognition	ACE-III ¹	95.57 (4.06) Range 66-100	95.95 (4.29)	94.99 (4.24)	96.12 (3.19)	Low: -0.17, 0.52, 0.74 Medium: -1.16, 0.49, 0.02	Low: 0.29, 0.77, 0.70 Medium: -0.88, 0.52, 0.09
Attention	Visual and auditory attention ²	9.82 (0.45) Range 7-10	9.82 (0.48)	9.82 (0.47)	9.85 (0.37)	Low: -0.04, 0.05, 0.43 Medium: -0.04, 0.04, 0.43	Low: -0.29, 0.98, 0.77 Medium: -0.16, 0.49, 0.74
Memory	Face-Name Recognition ³	5.26 (2.11) Range 0-9	5.20 (2.20)	5.19 (2.15)	5.50 (1.94)	Low: -0.27, 0.22, 0.21 Medium: -0.32, 0.21, 0.13	Low: -0.77, 4.45, 0.86 Medium: -0.56, 2.23, 0.80
	Implicit Memory ³	1.03 (0.67) Range -2.6-5.6	1.02 (0.64)	1.04 (0.67)	1.00 (0.73)	Low: 0.02, 0.07, 0.83 Medium: 0.04, 0.07, 0.57	Low: 2.35, 1.48, 0.11 Medium: 1.22, 0.74, 0.10
Visuospatial abilities	Geometric Forms ³	6.36 (1.17) Range 1-8	6.35 (1.11)	6.29 (1.20)	6.52 (1.18)	Low: -0.18, 0.12, 0.15 Medium: -0.23, 0.11, 0.04	Low: 2.93, 2.56, 0.25 Medium: 1.32, 1.29, 0.31
Language	Phoneme Comprehension ³	8.62 (0.58) Range 6-9	8.69 (0.57)	8.61 (0.60)	8.55 (0.58)	Low: 0.13, 0.06, 0.03 Medium: 0.06, 0.06, 0.34	Low: 1.54, 1.28, 0.23 Medium: 0.75, 0.64, 0.25
	Verbal Fluency (semantic) ⁴	16.36 (4.16) Range 0-29	15.99 (3.94)	16.56 (4.33)	16.54 (4.16)	Low: -0.55, 0.43, 0.20 Medium: 0.02, 0.41, 0.97	Low: -1.43, 8.49, 0.87 Medium: -0.39, 4.26, 0.93
	Verbal Fluency (phonemic) ³	11.26 (4.10) Range 1-24	11.21 (4.05)	11.13 (4.14)	11.60 (4.14)	Low: -0.39, 0.43, 0.26 Medium: -0.48, 0.40, 0.24	Low: 10.45, 8.98, 0.24 Medium: 4.93, 4.51, 0.27
Egocentric spatial orientation	VST ⁵	10.48 (2.19) Range 1-12	10.62 (2.25)	10.48 (2.17)	10.37 (2.19)	Low: 0.23, 0.28, 0.41 Medium: 0.11, 0.26, 0.68	Low: -4.71, 4.82, 0.33 Medium: 2.36, 2.42, 0.33
Allocentric spatial orientation	FMT ⁶	10.36 (2.41) Range 0-15	10.31 (2.34)	10.39 (2.50)	10.34 (2.41)	Low: -0.03, 0.31, 0.92 Medium: 0.05, 0.29, 0.86	Low: -1.46, 5.29, 0.78 Medium: -0.69, 2.66, 0.80

Table 6. Table of cognitive scores in total cohort and by risk group.

¹Total: n=467, low risk: n=151, medium risk: n=205, high risk: n=105). ²Total: n=691, low risk: n=228, medium risk: n=301, high risk: n=156. ³Total: n=693, low risk: n=228, medium risk: n=302, high risk: n=157. ⁴Total: n=692, low risk: n=228, medium risk: n=302, high risk: n=156. ⁵Total: n=453, low risk: n=141, medium risk: n=202, high risk: n=105. ⁶Total: n=459, low risk: n=147, medium risk: n=201, high risk: n=105. ACE-III: Addenbrookes Cognitive Examination III, FMT: Four Mountains Test, VST: Virtual Supermarket Trolley. Fully adjusted model adjusted for: sex, age, number of years of education, parental history of dementia and APOE ϵ 4.

Domain	Cognitive test	Female (mean (SD))	Male (mean (SD))	Unadjusted model β , SE, p (Comparator: Male)	Fully Adjusted model β , SE, p (Comparator: Male)
General cognition	ACE-III ¹	96.17 (3.87)	94.73 (4.17)	-1.44, 0.38, 0.0001	-1.32, 0.38, 0.0005
Attention	Visual and auditory attention ²	9.78 (0.51)	9.90 (0.33)	0.12, 0.04, 0.0009	0.13, 0.04, 0.0004
Memory	Face -Name Recognition ³	5.69 (2.02)	4.57 (2.08)	-1.13, 0.16, <0.001	-1.05, 0.16, <0.001
	Implicit Memory ³	1.02 (0.69)	1.04 (0.65)	0.03, 0.05, 0.59	0.03, 0.05, 0.63
Visuospatial abilities	Geometric Forms ³	6.37 (1.20)	6.35 (1.11)	-0.02, 0.09, 0.81	-0.004, 0.09, 0.97
Language	Phoneme Comprehension ³	8.65 (0.57)	8.58 (0.60)	-0.06, 0.05, 0.17	-0.06, 0.05, 0.20
	Verbal Fluency (semantic) ⁴	17.44 (3.97)	14.68 (3.88)	-2.76, 0.31, <0.001	-2.63, 0.30, <0.001
	Verbal Fluency (phonemic) ³	11.42 (4.11)	11.00 (4.08)	-0.42, 0.32, 0.19	-0.33, 0.32, 0.31
Egocentric spatial orientation	VST ⁵	10.12 (2.39)	10.97 (1.79)	0.85, 0.20, <0.001	0.90, 0.21, <0.001
Allocentric spatial orientation	FMT ⁶	10.28 (2.38)	10.46 (2.46)	0.18, 0.23, 0.43	0.29, 0.23, 0.21

Table 7. Table of cognitive scores by sex.

¹Female: n=267, male: n=197. ²Female: n=428, male: n=263. ³Female: n=428, male: n=265. ⁴Female: n=427, male: n=265. ⁵Female: n=260, male: n=193. ⁶Female: n=262, male: n=197. ACE-III: Addenbrookes Cognitive Examination III, FMT: Four Mountains Test, VST: Virtual Supermarket Trolley. Fully adjusted model adjusted for: age, years of education, parental history of dementia and *APOE* ϵ 4.

Imaging overview

From the completed 666 scans, 17 were excluded from analyses due to incidental findings (e.g., meningiomas) or poor quality of the imaging data. Our sample had an average WMH

volume of 2.26 ± 2.77 ml ($n=643$; median = 1.39 ml). This was higher than the mean of 0.95 ml in another midlife cohort of participants with mean age 45 [49]. However, this was expected given that our sample was older (mean age 51.2 years) and enriched for family history of dementia which may explain the high prevalence of *APOE* ϵ 4 carriers (37.7%) compared to the expected population prevalence of 20%. A small proportion (6.9%; $n=45$ out of 647) had a high burden of WMH, as defined by a Fazekas score of 3 in the periventricular area or a score of 2 in the deep subcortical white matter [50]. WMH volume did not differ by *APOE* ϵ 4 status or family history of dementia in unadjusted analysis or adjusted analyses controlling for sex, age, education, and site [18]. WMH burden increased with older age in both the unadjusted ($t = 5.73$, $p < 0.001$) and adjusted analysis controlling for sex, education, and site ($t = 5.40$, $p < 0.001$). Males (2.99 ml) had greater WMH volumes than females (1.81 ml), even after normalising for head size [(WMH volume/intracranial volume)*100%; males = 0.16, females = 0.11] – results were significant in both unadjusted ($\rho = 0.25$, $p < .001$) and adjusted analyses of WMH burden normalised by head size ($t = 8.88$, $p < .001$).

Following analysis with the FreeSurfer software (version 7.1.0), 623 datasets were free of incidental findings and artifacts and with good quality data following implementation of the *recon-all* pipeline. The mean hippocampal volume for the cohort (left and right hemispheres) was 8.15 ± 0.79 ml with an estimated total intracranial volume (eTIV) of 1490.6 ± 163.1 ml, gray matter volume of 646.3 ± 58.0 ml and cerebral white matter volume of 466.5 ± 56.8 ml. Mean cortical thickness was 2.43 ± 0.07 mm. In linear regression analysis with age, sex, education years, study site, eTIV and *APOE* ϵ 4 as predictors of hippocampal volume, sex was a significant predictor with females having smaller volumes ($t_{\text{female}} = -3.03$, $p < 0.01$). In a similar model predicting total GM volume, age ($t = -4.84$, $p < 0.01$), sex ($t_{\text{female}} = -10.44$, $p < 0.01$) and education years ($t = 2.24$, $p = 0.03$) were all significant predictors. Finally, mean cortical thickness was predicted by age ($t = -4.73$, $p < 0.01$) and years of education ($t = 2.19$, $p = 0.03$) [19]. Further description of the cohort imaging findings will be presented in an upcoming manuscript.

Prevalence of risk factors for Alzheimer's disease

Below we report the prevalence of common risk factors for AD as defined by the 2020 Lancet Commission on dementia prevention [51] as well as sleep as an important risk factor for brain health:

Early life:

- **Lower education:** 11.4% had less than 13 years of education. This is relatively consistent across all three risk groups (low-risk: 10.3%; medium-risk: 11.8%; high-risk: 12.1%) and both male and female participants (male: 11.9%; female: 11.1%).

Mid-life:

- **Hearing loss:** About 1 in 10 participants (11%) reported hearing loss during medical history taking. Participants with a medium-risk for future dementia had the highest rates of reported hearing loss, with low- and high-risk groups comparable (low-risk: 11.2%; medium-risk: 14.6%; high-risk: 10.8%). Male participants reported more hearing loss than female participants (female: 8%; male: 13.9%).
- **Head injury:** On the BISQ most participants (76.6%) reported ever receiving a blow to the head, while 28.2% reported five or more. Of those reporting at least one blow to the head, the average number of blows reported was 5.0 (SD = 5.5, range = 1 to 52), with 35.6% reporting at least one blow to the head resulting in a loss of consciousness and 50.9% reporting a feeling of being dazed and confused after a blow to the head. Males reported higher rates of TBI – half of males (49.6%) reported experiencing a TBI-LOC event, compared to 27.4% of females. Participants at lower risk were more likely to report a TBI-LOC event (low-risk: 37.5%; medium-risk: 36.9%, high-risk: 31.6%).
- **Hypertension:** 16.7% had Stage II hypertension, defined as mean systolic blood pressure ≥ 140 mmHg and/or mean diastolic blood pressure ≥ 90 mmHg. The prevalence of stage II hypertension was equivalent across all three risk groups (low-risk: 16.4%; medium-risk: 17.0%; high-risk: 16.6%). A higher proportion of male participants had stage II hypertension compared to female participants (male: 28.1%; female: 9.7%). A small number of the participants were taking medication for hypertension (7.7%), which was similar across risk groups (low-

risk: 7.8%; medium-risk: 8.2%; high-risk: 6.4%). Male participants had a higher prevalence of antihypertensive medication compared to female participants (male: 11.6%; female: 5.3%).

- **Excessive alcohol intake:** Nearly one-quarter of participants (24.8%) consumed >14 units of alcohol per week, while 14% consumed >21 units per week. Excessive alcohol intake was about twice as prevalent in males (>14 units per week: 35.4%; >21 units per week: 21.3%) than females (>14 units per week: 18.4%; >21 units per week: 9.4%). This did not differ by risk groups (>14 units per week: low-risk: 25.0%, medium-risk: 26.0%, high-risk: 23.2%; >21 units per week: low-risk: 14.2%, medium-risk: 14.2%, high-risk: 13.8%).
- **Obesity:** Based on body mass index (BMI), 4 in 10 (39.5%) participants fell into the overweight category (BMI 25-30), while one-quarter were obese (BMI ≥ 30), with similar prevalence for all risk groups (overweight low-risk: 41.8%; medium-risk: 38.0%; high-risk: 39.5%; obese low-risk: 28.0%; medium-risk: 25.9%; high-risk: 28.7%). A higher proportion of male participants met the criteria for being overweight compared to female participants (male: 51.7%; female: 32.1%), with no differences in proportions of obesity by sex (male: 28.5%; female: 26.3%). A small proportion were under-weight (1.1%), which has also been associated with increased risk of dementia. This did not differ by risk group (low-risk: 1.3%; medium-risk: 1.0%; high-risk: 1.3%), but there was a higher proportion of underweight females compared to males (female: 1.6%; male: 0.4%).

Later-life

- **Smoking:** Most participants were either non-smokers (58.2%), or ex-smokers (35.7%), while a small proportion were current smokers (5.6%). The proportion of smokers was lowest in the high-risk group (low-risk: 5.2%; medium-risk: 7.5%; high-risk: 2.5%), with more male participants reporting current smoking than female participants (male: 7.1%; female: 4.6%).
- **Depression:** 1 in 6 (16.7%) met the criterion for depression based on a cut-off of ≥ 16 on the (CES-D). The highest proportion of depression was seen in the low-risk group, and the lowest prevalence in the medium-risk group (low-risk: 20.7%; medium-risk: 13.4%; high-risk: 17.2%). There were equal proportions of

depression by sex (male: 16.9%; female: 16.6%). 8% of the cohort reported taking anti-depressant medication at the time of the baseline visit, with the highest proportion reported in the medium-risk group (low-risk: 6.5%; medium-risk: 8.9%; high-risk: 7.6%) and no difference by sex (male: 7.5%; female: 8.3%).

- **Social isolation:** One-quarter (26.8%) saw their friends and family less than daily, while 8.5% saw their friends and family less than once a week. This was more common in lower risk groups (less than daily – low-risk: 30.7%, medium-risk: 25.8%, high-risk: 22.6%; less than weekly – low-risk: 12.1%, medium-risk: 8.3%, high-risk: 3.1%), but did not differ by sex (less than daily – male: 27.3%, female: 26.7%; less than weekly – male: 9.7%, female: 7.9%).
- **Physical inactivity:** About half of participants (47.4%) were physically inactive, as defined as engaging in vigorous physical activity (e.g., running, tennis, digging) less than once a week AND moderately energetic activities (e.g., cycling, golf, lawn mowing) less than once a day. Physical inactivity was more common in females (53.4%) than males (38.6%), but did not differ by risk level (low-risk: 48.5%, medium-risk: 48.2%, high-risk: 46.5%). About half of participants (51.9%) engaged in vigorous physical activities less than once a week, while 36.5% did so less than once a month or never.
- **Diabetes mellitus:** A handful of participants reported a history of diabetes mellitus (3.3%). Proportions were comparable by risk group (low-risk: 3.9%; medium-risk: 3.0%; high-risk: 3.2%), but higher in male participants compared to female participants (male: 4.1%; female: 2.8%). Few participants were taking medication for diabetes (2.4%), which was comparable across the risk groups (low-risk: 3.0%; medium-risk: 2.0%; high-risk: 2.5%), with a higher proportion in male participants (male: 3.3%; female: 1.8%).
- **Sleep:** Average sleep was poor in the cohort, with a mean score on the PSQI of 6 (3.29). In total 315 participants (45%) met criteria for poor sleep using Buysse scoring methodology and cut off criteria [32]. Nearly half of the female participants reported poor sleep (47.1%) compared to 41.6% of male participants. The highest reports of poor sleep were in the low-risk group (47.8%), with the

medium- and high-risk groups reporting comparable proportions of poor sleep (medium-risk: 43.3%; high-risk: 47.8%).

Four participants had missing data on smoking, diabetes mellitus, and head injury. Three participants had missing data on BMI, hearing loss, education, social isolation, alcohol intake. Two participants had missing data on hypertension and physical inactivity. There are no data on air pollution exposure currently calculated for the cohort.

Discussion

The PREVENT dementia programme is a multi-site study with a comprehensive and deeply phenotyped baseline dataset from 700 participants recruited in midlife, an estimated 24 years from estimated dementia onset based on parental age of dementia onset. Data are available across a number of key early neurodegenerative disease indicators and risk factors for future neurodegenerative disease. Importantly this data collection has been collaboratively designed with an engaged participant panel. PREVENT participants are generally young and cognitively healthy. However, of importance to the field of dementia prevention, risk factors are already beginning to accumulate in this group. Of note, three-quarters (76.6%) of the cohort reported at least one head injury, 64.5% were overweight or obese, 47.4% were physically inactive and 45% had poor sleep. Male participants were carrying more of this burden, with higher rates of hearing loss, hypertension, being overweight, current smoking, TBI, alcohol use and diabetes. This midlife accumulation of risk factors highlights the importance of studying the origins of neurodegenerative disease in this age group. In fact, emerging evidence suggests that risk factors confer differential effects on brain health across the lifespan, whereby various risk factors are more predictive when measured at midlife, relative to late-life [52-55].

Given the early age and minimal cerebrovascular burden in the PREVENT cohort, it is well suited to delineate some of the earliest changes associated with risk factors of *APOE*ε4 and family history while mitigating risks of confounds from co-morbidity. The cognitive data presented in the manuscript showed no significant difference by a-priori risk groupings, but did suggest a number of sex differences in cognitive performance. As this manuscript was designed to be descriptive rather than hypothesis driven, the analysis were not designed to test any hypothesis regarding sex differences in midlife cognition, however the findings suggest further research in this topic is warranted, particularly given the emerging evidence in sex differences in the accumulation of AD pathology [56].

The collaborative core of PREVENT both with established cohorts (such as ALFA) and onboarding new sub-studies allows for both replication efforts and enrichment of the cohort. In particular, some of the sub-studies will provide data to support profiling of PREVENT participants using the Amyloid-Tau-Neurodegeneration (ATN) criteria as well as

analysis of stored blood using recently developed assays. These developments will allow researchers to study interactions between these pathological AD hallmarks with *APOE* ϵ 4 and family history of dementia. There is also opportunity to address questions that have not received much attention in the literature to date. For example can data from the PREVENT cohort help us to understand whether parental subtype of dementia is consequential, and whether AD-type parental dementia is associated with more deleterious outcomes vs non-AD parental dementia?

There are some notable limitations to the PREVENT cohort, namely around representative diversity. Particularly there is a lack of diversity in the ethnicity of participants, with the majority identifying as Caucasian, which has implications for both genetic analysis and the generalisability of findings to the UK and global populations. The cohort is also comparatively higher educated than the general adult population in the UK, which may limit generalisability of results to all groups of society.

The true potential of PREVENT is likely to be realised through both the release of the baseline data to the wider scientific community through the ADDI platform, and continued data collection. Additional and ongoing longer term follow up will also be beneficial to explore the symptomatic consequences of early pathological disease accumulation.

Data Availability

The baseline dataset is available to access through a data request on the study website (www.preventdementia.co.uk); on the Alzheimer's Disease Data Initiative (ADDI) platform baseline dataset DOI: https://doi.org/10.34688/PREVENTMAIN_BASELINE_700V1; Dementia Platforms UK (DPUK); and the Global Alzheimer's Association Network (GAAIN).

For imaging data a number of derived variables (for example the global volumetrics and WMH volume) are available in the ADDI dataset, with raw structural data available to access upon request following defacing.

Acknowledgements

We would like to acknowledge the sites involved with the project, West London NHS Trust, NHS Lothian, Cambridgeshire and Peterborough NHS Foundation trust, Oxford Heath NHS Foundation trust and Trinity College. Special thanks also to the PREVENT participants, the participant panel, members of the Scientific Advisory Committee, and funders for their support of the PREVENT dementia programme.

Author contributions

CWR is the principal investigator of the protocol and oversaw the design and statistical analysis and worked on the first draft of the manuscript. KW, SG, AL, MED, EM and RH all drafted the manuscript and provided comments on subsequent drafts. SG and AL additionally completed the statistical analysis for the paper. SD provided support for the statistical analysis and provided comments on the manuscript drafts. JOB, SL, WS and GMT all contributed to the design of the protocol and provided comments on the manuscript drafts. KR is the co-principal investigator and oversaw the design, and worked on the first draft of the manuscript

Funding

PREVENT is funded by the Alzheimer's Society (grant numbers 178, 264 and 329), Alzheimer's Association (grant number TriBEKa-17-519007) and philanthropic donations. Sub-studies have their own funding sources which are not detailed here. GMT acknowledges the support of the Osteopathic Heritage Foundation through funding for the Osteopathic Heritage Foundation Ralph S. Licklider, D.O., Research Endowment in the Heritage College of Osteopathic Medicine. IK declares support for the work on this project through the Oxford Health Biomedical Research Centre, Medical Research Council Dementias Platform UK project as well as personal awards (NIHR Academic Lectureship and NIHR Development and Skills Enhancement award).

Competing interests

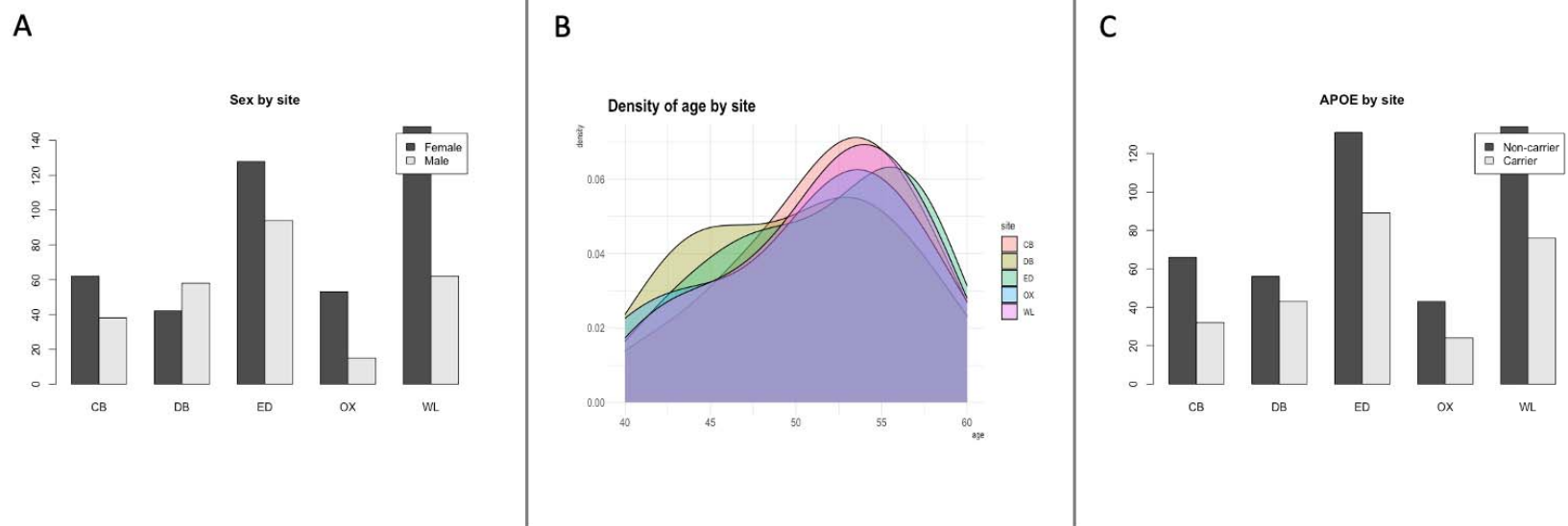
The authors report no competing interests.

References

1. Ritchie, C.W. and K. Ritchie, *The PREVENT study: a prospective cohort study to identify mid-life biomarkers of late-onset Alzheimer's disease*. BMJ Open 2012. **2**(e001893).
2. Ritchie, C.W., K. Wells, and K. Ritchie, *The PREVENT research programme--a novel research programme to identify and manage midlife risk for dementia: the conceptual framework*. Int Rev Psychiatry, 2013. **25**(6): p. 748-54.
3. Molinuevo, J.L., et al., *The ALFA project: A research platform to identify early pathophysiological features of Alzheimer's disease*. Alzheimer's & Dementia: Translational Research & Clinical Interventions, 2016. **2**(2): p. 82-92.
4. Ritchie, C.W., et al., *Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project*. Lancet Psychiatry, 2016. **3**(2): p. 179-186.
5. Ritchie, C.W., et al., *The European Prevention of Alzheimer's Dementia (EPAD) Longitudinal Cohort Study: Baseline Data Release V500.0*. The Journal of Prevention of Alzheimer's Disease, 2020. **7**(1): p. 8-13.
6. Solomon, A., et al., *European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol*. BMJ Open, 2018. **8**(12): p. e021017.
7. Gregory, S., et al., *Involving research participants in a pan-European research initiative: the EPAD participant panel experience*. Research Involvement and Engagement, 2020. **6**(1): p. 62.
8. Milne, R., et al., *At, with and beyond risk: expectations of living with the possibility of future dementia*. Sociol Health Illn, 2018. **40**(6): p. 969-987.
9. Milne, R., et al., *Ethical Issues in the Development of Readiness Cohorts in Alzheimer's Disease Research*. J Prev Alzheimers Dis, 2017. **4**(2): p. 125-131.
10. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets*. GigaScience, 2015. **4**(1): p. s13742-015-0047-8.
11. Das, S., et al., *Next-generation genotype imputation service and methods*. Nature Genetics, 2016. **48**(10): p. 1284-1287.
12. Taliun, D., et al., *Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program*. Nature, 2021. **590**(7845): p. 290-299.
13. Fuchsberger, C., G.R. Abecasis, and D.A. Hinds, *minimac2: faster genotype imputation*. Bioinformatics, 2015. **31**(5): p. 782-784.
14. Fischl, B., *FreeSurfer*. NeuroImage, 2012. **62**(2): p. 774-781.
15. Tustison, N.J., et al., *N4ITK: improved N3 bias correction*. IEEE Trans Med Imaging, 2010. **29**(6): p. 1310-20.
16. Firbank, M.J., T. Minett, and J.T. O'Brien, *Changes in DWI and MRS associated with white matter hyperintensities in elderly subjects*. Neurology, 2003. **61**(7): p. 950-4.
17. Low, A., et al., *Inherited risk of dementia and the progression of cerebral small vessel disease and inflammatory markers in cognitively healthy midlife adults: the PREVENT-Dementia study*. Neurobiol Aging, 2021. **98**: p. 124-133.
18. Low, A., et al., *Modifiable and non-modifiable risk factors of dementia on midlife cerebral small vessel disease in cognitively healthy middle-aged adults: the PREVENT-Dementia study*. Alzheimer's Research & Therapy, 2022. **14**(1): p. 154.

19. Dounavi, M.-E., et al., *Macrostructural brain alterations at midlife are connected to cardiovascular and not inherited risk of future dementia: the PREVENT-Dementia study*. Journal of Neurology, 2022. **269**(8): p. 4299-4309.
20. Nili, H., et al., *A toolbox for representational similarity analysis*. PLOS Computational Biology, 2014. **10**(4): p. e1003553.
21. Low, A., et al., *CAIDE dementia risk score relates to severity and progression of cerebral small vessel disease in healthy midlife adults: the PREVENT-Dementia study*. Journal of Neurology, Neurosurgery & Psychiatry, 2022. **93**(5): p. 481.
22. Ritchie, K., et al., *COGNITO: Computerized Assessment of Information Processing*. Journal of Psychology and Psychotherapy, 2014. **4**(2).
23. Ritchie, K., et al., *Allocentric and Egocentric Spatial Processing in Middle-Aged Adults at High Risk of Late-Onset Alzheimer's Disease: The PREVENT Dementia Study*. Journal of Alzheimer's Disease, 2018. **65**: p. 885-896.
24. Nelson, H.E. and J. Willison, *National Adult Reading Test (NART)*. 1991: Windsor: NFER-Nelson.
25. Tu, S., et al., *Lost in spatial translation - A novel tool to objectively assess spatial disorientation in Alzheimer's disease and frontotemporal dementia*. Cortex, 2015. **67**: p. 83-94.
26. Parra, M.A., et al., *Visual short-term memory binding deficits in familial Alzheimer's disease*. Brain, 2010. **133**(9): p. 2702-13.
27. Noone, P., *Addenbrooke's Cognitive Examination-III*. Occupational Medicine, 2015. **65**(5): p. 418-420.
28. Valenzuela, M.J. and P. Sachdev, *Assessment of complex mental activity across the lifespan: development of the Lifetime of Experiences Questionnaire (LEQ)*. Psychol Med, 2007. **37**(7): p. 1015-25.
29. International Physical Activity Questionnaire. 2002; Available from: https://sites.google.com/site/theipaq/questionnaire_links.
30. Radloff, L.S., *The CES-D Scale: A self-report depression scale for research in the general population*. 1977, Sage Publications: US. p. 385-401.
31. Spielberger, C.D., et al., *Manual for the State-Trait Anxiety Inventory*. . 1983, Consulting Psychologists Press: Palo Alto, CA.
32. Buysse, D.J., et al., *The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research*. Psychiatry Res, 1989. **28**(2): p. 193-213.
33. Johns, M.W., *The Epworth Sleepiness Questionnaire*. 1990.
34. Connor, K.M. and J.R. Davidson, *Connor-Davidson Resilience scale (CD-RISC)*. . 2001, Duke University Medical Center: Durham, North Carolina.
35. Wolfe, J., et al., *Life Stressor Checklist- Review (LSC-R)*. . 1996, APA PsycTests.
36. *Brain Injury Screening Questionnaire* 2011, Brain Injury Research Center, Department of Rehabilitation Medicine, Mount Sinai Medical Center: New York.
37. Scottish Collaborative Group, *Scottish collaborative group food frequency questionnaire*. 2004.
38. de Jong, F.J., et al., *Retinal vascular caliber and risk of dementia: the Rotterdam study*. Neurology, 2011. **76**(9): p. 816-21.
39. Frost, S., et al., *Retinal vascular biomarkers for early detection and monitoring of Alzheimer's disease*. Transl Psychiatry, 2013. **3**(2): p. e233.
40. Cheung, C.Y., et al., *Microvascular network alterations in the retina of patients with Alzheimer's disease*. Alzheimers Dement, 2014. **10**(2): p. 135-42.
41. Williams, M.A., et al., *Retinal microvascular network attenuation in Alzheimer's disease*. Alzheimers Dement (Amst), 2015. **1**(2): p. 229-235.
42. Ukalovic, K., et al., *Drusen in the Peripheral Retina of the Alzheimer's Eye*. Curr Alzheimer Res, 2018. **15**(8): p. 743-750.

43. Milne, R., et al., *Perspectives on Communicating Biomarker-Based Assessments of Alzheimer's Disease to Cognitively Healthy Individuals*. J Alzheimers Dis, 2018. **62**(2): p. 487-498.
44. de la Fuente Garcia, S., C.W. Ritchie, and S. Luz, *Protocol for a conversation-based analysis study: PREVENT-ED investigates dialogue features that may help predict dementia onset in later life*. BMJ Open, 2019. **9**(3): p. e026254.
45. Gregory, S., et al., *Research participants as collaborators: Background, experience and policies from the PREVENT Dementia and EPAD programmes*. Dementia (London), 2018. **17**(8): p. 1045-1054.
46. Harris, P.A., et al., *The REDCap consortium: Building an international community of software platform partners*. J Biomed Inform, 2019. **95**: p. 103208.
47. González, D.R., et al., *An open source toolkit for medical imaging de-identification*. Eur Radiol, 2010. **20**(8): p. 1896-904.
48. Hsieh, S., et al., *Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal Dementia and Alzheimer's Disease*. Dementia and Geriatric Cognitive Disorders, 2013. **36**(3-4): p. 242-250.
49. d'Arbeloff, T., et al., *White matter hyperintensities are common in midlife and already associated with cognitive decline*. Brain Commun, 2019. **1**(1): p. fcz041.
50. Fazekas, F., et al., *MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging*. AJR Am J Roentgenol, 1987. **149**(2): p. 351-6.
51. Livingston, G., et al., *Dementia prevention, intervention, and care: 2020 report of the Lancet Commission*. The Lancet, 2020. **396**(10248): p. 413-446.
52. Debette, S., et al., *Midlife vascular risk factor exposure accelerates structural brain aging and cognitive decline*. Neurology, 2011. **77**(5): p. 461-8.
53. Tolppanen, A.M., et al., *Midlife and late-life body mass index and late-life dementia: results from a prospective population-based cohort*. J Alzheimers Dis, 2014. **38**(1): p. 201-9.
54. Dickie, D.A., et al., *Vascular risk factors and progression of white matter hyperintensities in the Lothian Birth Cohort 1936*. Neurobiol Aging, 2016. **42**: p. 116-23.
55. Jorgensen, D.R., et al., *A population neuroscience approach to the study of cerebral small vessel disease in midlife and late life: an invited review*. American Journal of Physiology-Heart and Circulatory Physiology, 2018. **314**(6): p. H1117-H1136.
56. Saunders, T.S., et al., *Interactions between apolipoprotein E, sex, and amyloid-beta on cerebrospinal fluid p-tau levels in the European prevention of Alzheimer's dementia longitudinal cohort study (EPAD LCS)*. eBioMedicine, 2022. **83**.



Supplementary Figure S1: Distribution of sex (A), age (B) and *APOE* ϵ 4 (C) by site.

		Sample Size	PCT_1%	PCT_5%	PCT_10%	PCT_25% Q1	PCT_50% Median	PCT_75% Q3	PCT_90%	PCT_95%	PCT_99%
EP01M	Task 1: reaction time, mean (millisec)	695	262	277	287	306	328	359	388	410	455
EP03BR	Task 3: sentence comprehension, No correct answers	700	3	4	4	5	5	5	5	5	5
EP04BRT	Task 4: Auditory attention, total number of correct counts (max=10)	699	7	8	9	10	10	10	10	10	10
EP05BRT	Task 5: Total correct answers (max=10=5 trials with 2 forms)	700	7	9	9	10	10	10	10	10	10
EP05T1M	Task 5: Auditory attention, mean time (millisec) until 1st click (whatever the response)	700	1687	2067	2248.5	2647.5	3259.5	3994.5	4909	5541	7363.5
EP06BRT	Task 6: Total number of correct answers for form recognition (max=10=5*2)	698	8	9	9	10	10	10	10	10	10
EP06T1M	Task 6: Working memory, mean time (millisec) until 1st click (whatever the response)	698	1623	1948	2130	2513	3113	4022	4973	6233	8900
EP07ET02YT	Task 7, level 2: Mean duration of answers (milliseconds)	696	686	749	791	856	941	1029	1143	1213	1422
EP07ET03BR	Task 7, level 3: Stroop proactive inference, number of correct answers	696	4	13	15	17	18	20	21	22	23
EP08BR	Task 8: Immediate name recall, number of names correctly recalled	699	4	4	5	6	7	8	8	9	9
EP10RE	Task 10: Spatial span, number of successes	699	0	2	3	3	5	5	6	6	6
EP10T	Task 10: Spatial span, mean time (milliseconds)	699	880	1303	1420	1764	2234.2	2918.5	3605.8	4283.5	5570.5
EP11BR	Task 11, geometric forms: No correct responses	700	3	4	5	6	7	7	8	8	8
EP11MY	Task 11, Mean time for correct responses (milliseconds)	700	3101.5	3629.5	3973.5	4740	5592.5	6608.5	7839	8663	9871
EP12BR	Task 12: Phoneme comprehension, No of correct answers (max=10)	700	7	7	8	8	9	9	9	9	9
EP12MY	Task 12, word comprehension: mean time for correct response (milliseconds)	700	943	1097.5	1151.5	1273	1420	1617	1834.5	1955	2272
EP13BA_P	Task 13: Semantic associations, No correct associations(max=10)	700	8	9	9	9	10	10	10	10	10
EP13RDP	Task 13: Semantic associations, No trials over time limit	700	0	0	0	0	0	0	0	0	0
EP13RI	Task 13: Semantic associations, No of exact responses for naming	700	8	9	10	10	10	10	10	10	10
EP14BR_P	Task 14: Visuospatial logic (matrices), nb of correct answers	699	3	6	7	8	9	11	12	13	14
EP14TT	Task 14: Visuospatial logic (matrices), Total time	699	84387	104287	114612	139216	177955	229689	278682	309734	371314
EP17ET01BR	Task 17, Delayed name recall, free recall: no of correct	699	3	4	5	6	7	8	9	9	9

EP17ET02BR	answers Task 17, Delayed name recall, cued recall: no of correct answers	700	3	4	5	6	7	8	9	9
EP18BN	Task 18, surname recognition: no. of names correctly recognized	700	1	2	2	4	5	7	8	9
EP19_20C	Tasks 19 and 20, verbal fluency, semantic phonemic cue: nb of correct answers 60s	699	14	17	19	23	28	32	37	39
EP21T	Task 21, narrative story: total nb of correct answers (max=27)	700	3	7	8	11	14	17	20	21
EP22T	Task 22, descriptive story: total nb of correct answers (max=27)	700	5	6.5	8	10	13	17	19.5	21
EP23BR	Task 23: Nb of correct answers (max=35)	700	8	12	14	18	21	25	28	29
EP23MR	Task 23: Vocabulary, no of incorrect answers	700	3	6	7	10	13	17	21	23
EP24DF	Task 24 implicit memory: difference	700	-0.4	0	0.2	0.6	1	1.4	1.8	2.1
EP25T_P	Task 25: scoring the design	607	45	55	62	68	73	77	80	80

Supplementary Table 1: Percentiles of COGNITO tasks in all participants.

		Sex	Education	Sample Size	PCT_1%	PCT_5%	PCT_10%	PCT_25% Q1	PCT_50% Median	PCT_75% Q3	PCT_90%	PCT_95%	PCT_99%
EP01M	Task 1: reaction time, mean (millsec)	Male	<18 years	175	262	276	284	303	322	348	369	394	431
		Male	18+	90	253	273	285.5	300	319.5	352	390	412	460
		Female	<18 years	255	262	280	297	315	342	374	398	425	444
		Female	18+	175	264	278	287	302	323	352	384	408	456
EP03BR	Task 3: sentence comprehension, No correct answers	Male	<18 years	176	3	4	4	5	5	5	5	5	5
		Male	18+	91	3	4	4	5	5	5	5	5	5
		Female	<18 years	256	2	3	4	5	5	5	5	5	5
		Female	18+	177	3	4	4	5	5	5	5	5	5
EP04BRT	Task 4: Auditory attention, total number of correct counts (max=10)	Male	<18 years	175	4	8	9	9	10	10	10	10	10
		Male	18+	91	7	9	9	10	10	10	10	10	10
		Female	<18 years	256	7	8	9	9	10	10	10	10	10
		Female	18+	177	7	8	9	10	10	10	10	10	10
EP05BRT	Task 5: Total correct answers (max=10=5 trials with 2 forms)	Male	<18 years	176	8	10	10	10	10	10	10	10	10
		Male	18+	91	8	9	10	10	10	10	10	10	10
		Female	<18 years	256	7	9	9	10	10	10	10	10	10
		Female	18+	177	5	9	9	10	10	10	10	10	10
EP05T1M	Task 5: Auditory attention, mean time (millisec) until 1st click (whatever the response)	Male	<18 years	176	1755	1953	2125	2514	3060	3696.5	4483	4980	7393
		Male	18+	91	1514	1972	2131	2509	3096	3851	4427	5182	6614
		Female	<18 years	256	1866	2246	2402	2785.5	3505	4312.5	5410	5604	7553
		Female	18+	177	1552	1951	2253	2600	3288	3859	4727	5557	7954
EP06BRT	Task 6: Total number of correct answers for form recognition (max=10=5*2)	Male	<18 years	174	8	9	10	10	10	10	10	10	10
		Male	18+	91	9	9	10	10	10	10	10	10	10
		Female	<18 years	256	8	9	9	10	10	10	10	10	10

EP06T1M	Task 6: Workin memory, mean time (millisec) until 1st click (whatever the response)	Female	18+	177	8	9	9	10	10	10	10	10	10
		Male	<18 years	174	1526	1948	2072	2458	2950.5	3561	4487	4841	6964
		Male	18+	91	1533	1792	2081	2337	2824	3738	4429	4666	12517
		Female	<18 years	256	1763	2047	2191	2665	3262	4534	5990	7340	9738
EP07ET02YT	Task 7, level 2: Mean duration of answers (milliseconds)	Female	18+	177	1598	1876	2123	2574	3271	4097	4756	5778	8457
		Male	<18 years	173	690	743	796	874	959	1025	1143	1228	1526
		Male	18+	90	673	727	785.5	849	927	1036	1145.5	1205	1345
		Female	<18 years	256	699	751	800	855.5	952	1051.5	1148	1259	1706
EP07ET03BR	Task 7, level 3: Stroop proactive inference, number of correct answers	Female	18+	177	686	759	777	850	922	1016	1102	1171	1259
		Male	<18 years	173	9	12	14	16	18	20	21	22	23
		Male	18+	90	11	15	16	17	18	20	21	22	23
		Female	<18 years	256	4	13	15	17	18	20	21	22	23
EP08BR	Task 8: Immediate name recall, number of names correctly recalled	Female	18+	177	4	13	15	17	19	20	21	22	23
		Male	<18 years	175	3	4	4	5	6	7	8	8	9
		Male	18+	91	4	4	5	5	6	7	8	8	9
		Female	<18 years	256	4	5	5	6	7	8	8	9	9
EP10RE	Task 10: Spatial span, number of successes	Female	18+	177	4	5	5	6	7	8	8	9	9
		Male	<18 years	175	0	2	3	3	5	6	6	6	6
		Male	18+	91	0	3	3	3	5	5	6	6	6
		Female	<18 years	256	0	2	3	3	4	5	6	6	6
EP10T	Task 10: Spatial span, mean time (milliseconds)	Female	18+	177	0	2	3	3	5	5	6	6	6
		Male	<18 years	175	880	1343.8	1422.5	1804	2224.8	2921.8	3704	4283.5	5571
		Male	18+	91	719.5	1375.8	1530.7	1694.8	2158.3	3096.2	3839.5	4918.7	6994.3
		Female	<18 years	256	909.8	1205.3	1412.7	1784.3	2300.5	2959.3	3486.2	4226	5509.6
EP11BR	Task 11, geometric forms: No correct	Female	18+	177	521	1293.8	1375.7	1707	2215.5	2678.5	3494.8	3931.2	5849.2
		Male	<18 years	176	3	4	5	6	7	7	7	8	8

EP11MY	Task 11, Mean time for correct responses (milliseconds)	responses											
		Male	18+	91	3	4	5	6	6	7	8	8	8
		Female	<18 years	256	2	4	5	6	6	7	8	8	8
		Female	18+	177	3	4	5	6	7	7	8	8	8
		Male	<18 years	176	3277	3629	3918	4702	5508.5	6629	7779	8703	9933
		Male	18+	91	3142	3287	3623	4470	5646	6613	7501	8376	11916
		Female	<18 years	256	2930	3638	4070	4825.5	5638.5	6793	7988	8945	10792
		Female	18+	177	3381	3732	4095	4735	5604	6315	7779	8303	9327
		Male	<18 years	176	7	7	8	8	9	9	9	9	9
EP12BR	Task 12: Phoneme comprehension, No of correct answers (max=10)	Male	18+	91	7	7	8	8	9	9	9	9	9
		Female	<18 years	256	7	8	8	8	9	9	9	9	9
		Female	18+	177	7	8	8	8	9	9	9	9	9
		Male	<18 years	176	829	1103	1157	1287	1433.5	1642	1913	2053	2452
EP12MY	Task 12, word comprehension: mean time for correct response (milliseconds)	Male	18+	91	955	1148	1218	1325	1452	1664	1874	2150	2769
		Female	<18 years	256	943	1087	1138	1273	1426	1644	1840	1940	2263
		Female	18+	177	970	1093	1126	1245	1379	1536	1733	1899	2198
		Male	<18 years	176	8	9	9	10	10	10	10	10	10
EP13BA_P	Task 13: Semantic associations, No correct associations(max=10)	Male	18+	91	8	9	9	9	10	10	10	10	10
		Female	<18 years	256	8	9	9	9	10	10	10	10	10
		Female	18+	177	8	9	9	9	10	10	10	10	10
		Male	<18 years	176	0	0	0	0	0	0	0	0	1
EP13RDP	Task 13: Semantic associations, No trials over time limit	Male	18+	91	0	0	0	0	0	0	0	0	1
		Female	<18 years	256	0	0	0	0	0	0	0	1	1
		Female	18+	177	0	0	0	0	0	0	0	0	1
		Male	<18 years	176	9	9	10	10	10	10	10	10	10
EP13RI	Task 13: Semantic associations, No of exact responses for naming	Male	18+	91	8	10	10	10	10	10	10	10	10
		Female	<18 years	256	8	9	9	10	10	10	10	10	10
		Male	<18 years	176	9	9	10	10	10	10	10	10	10

EP14BR_P	Task 14: Visiospatial logic (matrices), nb of correct answers	Female	18+	177	8	9	10	10	10	10	10	10	10
		Male	<18 years	175	2	6	7	8	9	11	12	13	14
		Male	18+	91	3	6	7	8	9	11	12	13	15
		Female	<18 years	256	3	5	6	7	9	10	11	12	13
		Female	18+	177	2	7	7	8	10	11	12	13	14
EP14TT	Task 14: Visiospatial logic (matrices), Total time	Male	<18 years	175	84387	104287	114961	145826	186308	217060	258094	283564	349296
		Male	18+	91	79550	106312	114869	134727	170101	207039	277068	327761	409062
		Female	<18 years	256	80731	98493	111836	133287.5	175731.5	236157	286123	306610	351042
		Female	18+	177	93197	106768	121270	142523	184105	238717	282815	309734	420742
		Male	<18 years	175	2	3	4	5	6	7	8	9	9
EP17ET01BR	Task 17, Delayed name recall, free recall: no of correct answers	Male	18+	91	3	4	5	6	7	8	8	8	9
		Female	<18 years	256	3	5	5	6	7	8	9	9	9
		Female	18+	177	3	5	6	7	8	8	9	9	9
		Male	<18 years	176	2	3	4	5	7	7	8	9	9
		Male	18+	91	4	5	5	6	7	8	8	9	9
EP17ET02BR	Task 17, Delayed name recall, cued recall: no of correct answers	Female	<18 years	256	3	5	6	6	7	8	9	9	9
		Female	18+	177	3	5	6	7	8	8	9	9	9
		Male	<18 years	176	0	1	2	3	4	6	7	8	9
		Male	18+	91	0	2	2	3	5	6	8	8	9
		Female	<18 years	256	1	2	3	4	6	7	8	9	9
EP18BN	Task 18, surname recognition: no. of names correctly recognized	Female	18+	177	1	2	3	5	6	7	8	9	9
		Male	<18 years	176	0	1	2	3	4	6	7	8	9
		Male	18+	91	0	2	2	3	5	6	8	8	9
		Female	<18 years	256	1	2	3	4	6	7	8	9	9
		Female	18+	177	1	2	3	5	6	7	8	9	9
EP19_20C	Tasks 19 and 20, verbal fluency, semantic phonemic cue: nb of correct answers 60s	Male	<18 years	176	14	16	17	20	24	29.5	33	37	40
		Male	18+	91	11	15	18	22	28	31	35	37	42
		Female	<18 years	255	14	17	19	24	28	32	36	39	42
		Female	18+	177	17	20	21	25	30	34	38	42	45
		Male	<18 years	176	1	6	8	11	14	17	19	21	23
EP21T	Task 21, narrative story: total nb of	Male	<18 years	176	1	6	8	11	14	17	19	21	23

	correct answers (max=27)												
EP22T	Task 22, descriptive story: total nb of correct answers (max=27)	Male	18+	91	5	7	8	11	13	17	20	22	24
		Female	<18 years	256	3	6	8	11	14	17	19	21	24
		Female	18+	177	5	8	9	13	15	18	20	21	26
		Male	<18 years	176	4	6	7	9	12	15	18	21	23
		Male	18+	91	0	6	8	10	14	16	18	21	24
		Female	<18 years	256	5	6	8	10	13	17	20	21	23
EP23BR	Task 23: Nb of correct answers (max=35)	Female	18+	177	5	9	10	12	15	17	21	22	24
		Male	<18 years	176	8	12	14	17	21	25	28	29	33
		Male	18+	91	7	13	15	20	24	26	28	31	33
		Female	<18 years	256	6	11	12	16	19	23.5	27	28	31
		Female	18+	177	11	14	17	20	24	26	29	30	32
		Male	<18 years	176	2	6	7	10	14	17	20	23	27
EP23MR	Task 23: Vocabulary, no of incorrect answers	Male	18+	91	2	4	7	9	11	14	19	22	26
		Female	<18 years	256	4	7	8	11	15	19	23	24	26
		Female	18+	177	3	5	6	9	11	15	18	20	23
		Male	<18 years	176	-0.2	0.2	0.2	0.6	1	1.4	1.8	2	3.4
		Male	18+	91	-0.2	0.2	0.4	0.6	1	1.4	1.8	2.4	2.6
		Female	<18 years	256	-0.4	0	0.2	0.6	1	1.4	1.8	2.2	3
EP24DF	Task 24 implicit memory: difference	Female	18+	177	-1	0	0.2	0.6	1	1.4	1.8	2	2.4
		Male	<18 years	146	47	55	61	68	73	77	80	80	80
		Male	18+	80	39	45.5	62.5	67	72.5	77	79.5	80	80
		Female	<18 years	232	46	55	62	68	72.5	77	80	80	80
		Female	18+	149	51	55	62	70	74	77	80	80	80
EP25T_P	Task 25: scoring the design												

Supplementary Table 2: Percentiles of COGNITO tasks by sex and education.