Nature Reviews Neurology
Final version 16 April 2025

UK Biobank at 20 years old: a growing, global resource for dementia research

Paul M. Matthews[1,2], Naomi E. Allen[3,4], Stephanie Debette[5,6,7], Aiden Doherty[3], Gwen Douaud[8], Eugene P. Duff[1], Paul Elliot,[9] Evelynne Fulda[3], Anastasia Illina[1], Quentin Le Grand[7], Adam J. Lewandowski[3], Rebecca Mahoney[10], Karla Miller[8], Cecilia Rodriguez[1], Martin K Rutter[11,12], Cynthia Sandor[1], Rebecca Sims[13], Stephen M. Smith[8] and Chaoyue Wang[8,14,15]

[1]UK Dementia Research Institute Centre and Department of Brain Sciences, Imperial College London, UK; [2]Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, UK; [3]Nuffield Department of Population Health, University of Oxford, Oxford UK; [4]UK Biobank Ltd., Stockport, UK; [5]Institut du Cerveau (ICM), Paris Brain Institute, INSERM U1127, UMR CNRS 7225 Paris, Sorbonne Université, Assistance Publique des Hôpitaux de Paris, Paris, France ; [6]Department of Neurology, Institute for Neurodegenerative Diseases, Bordeaux University Hospital, Bordeaux, France;[7]University of Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR1219, F-33000, Bordeaux, France; [8]University of Oxford Centre for Integrative Neuroimaging (OxCIN), Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK; [9]MRC Centre for Environment and Health and UK Dementia Research Institute Centre, School of Public Health, Imperial College London, UK; [10]UK Dementia Research Institute at Cardiff, Cardiff, UK; [11]Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK; [12]Diabetes, Endocrinology and Metabolism Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, UK; [13]Division of Psychological Medicine, School of Medicine, Cardiff University, UK.; [14]Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; [15]SJTU-Ruijin-UIH Institute for Medical Imaging Technology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

Corresponding author:
Prof. Paul Matthews
Dept of Brain Sciences
Uren Building, Floor 7
White City Campus
Imperial College
London Wt12 7TQ
Email: p.matthews@imperial.ac.uk

Abstract 129 words
Major points 136 words
Main text 4321 words
Boxes 869 words
Table 936 words
Figures 6

Acknowledgements:

PMM thanks the Edmond J Safra Foundation and Lily Safra for their long-term, generous support and gratefully acknowledges an NIHR Senior Investigator Award and infrastructure support from the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) research funding from the UK Dementia Research Institute, which receives funding from UK DRI Ltd., funded by the UK Medical Research Council, Alzheimer's Society, and Alzheimer's Research UK. SD acknowledges support from the French National Research Agency and France 2030 (ANR-18-RHUS-0002, RHU-SHIVA; ANR-23-IAHU-0001, IHU-VBHI), Prix Burrus-FRM and NRJ-neurosciences, EU Horizon 2020 (grant No 754517). SD and QLG acknowledge support from the Fondation Recherche Alzheimer. AD is supported by a Wellcome Trust Senior Research Fellowship [223100/Z/21/Z]. EPD is supported jointly by funding to PMM from the UK Dementia Research Institute and from the NIHR Imperial Biomedical Research Centre Multiple Long-Term Conditions theme.PE acknowledges support from the UK Dementia Research Institute, Health Data Research UK, Medical Research Council, National Institute for Health and Care Research (NIHR), and the NIHR Health Protection Research Unit in Radiation Threats and Hazard. EF is supported by the National Institutes of Health's Oxford Cambridge Scholars Program. CS, AI, and CR are supported by the UK Dementia Research Institute [award number UK DRI-5209] and a UKRI Future Leaders Fellowship [MR/X032892/1]. CS receives personal support from the Edmond J. Safra Foundation. RS received funding from Alzheimer's Society, Alzheimer's Research UK and BRACE.

Declarations:

PMM is Chair of the UK Biobank Imaging Working Group, a consultant for Biogen, Nodthera, Sudo Therapeutics, GSK and Novartis and has received research funding from Biogen, Merck and Bristol Meyers Squibb. NEA is Chief Scientist for UK Biobank and AJL and MKR are Deputy Chief Scientists for UK Biobank with funding from Wellcome, Medical Research Council, British Heart Foundation and Cancer Research UK. AD receives support through a range of grants from Novo Nordisk, Swiss Re, Boehringer Ingelheim, and GSK. PE is a member of the UK Biobank International Scientific Advisory Board.

Abstract:

UK Biobank, which was started 20 years ago, is the world's most comprehensive longitudinal population-based data and biosample resource. The data that has been collected can be accessed by all *bona fide* researchers. In addition to the extensive clinical phenotyping of the over 500,000 volunteers from across the UK at recruitment and at follow up timepoints, it includes data from serial lifestyle questionnaires, cognitive testing, multi-modal imaging, accelerometry, genomic and other 'omics and is linked to participants' electronic healthcare, cancer, death and a proportion of the cohort's primary care records. Here, with a focus on dementia, we provide examples of how UK Biobank data has allowed discovery of novel interactions between systemic and brain health and illustrate ways in which these data also can be used to characterise risk factors, support mechanistic hypotheses and discover new biomarkers predictive of the onset and course of dementia and related disorders.

Major points:

- UK Biobank, which includes almost two decades of follow up on its participants, is a major resource for neuroepidemiology that is open for use by all *bona fide* researchers.
- The longitudinal design, which allows many participants to be characterised before clinical disease expression, and the availability of genetic data for complementary Mendelian randomisation studies together support researchers for exploring causal relationships between risk factors and dementia.
- Major contributions to characterisation of modifiable lifestyle and exposure risk factors for dementia have been made using these data.
- Data resource enhancements for genetics, accelerometry, imaging, serological markers of infection exposures and blood 'omics are accelerating population-based studies of disease mechanisms and new biomarker discovery.
- The value of UK Biobank for dementia research will continue to grow as exposure and systemic health outcome data are progressively expanded, the population ages and numbers of incident cases of dementia increase.

Introduction

The growing burden of dementia is one of the defining medical challenges for this generation. Establishing evidence-based strategies for early specific diagnoses of dementia and their management or prevention has become an urgent goal. The promise of discovering links between dementia and other health conditions or environmental exposures has long been recognised. Research also has made clear that many dementias develop and progress for extended periods without clinical symptoms.[1,2] This has created the alluring promise of slowing or preventing dementias by modifying chronic risk factors or of developing new therapeutics based on emerging understanding of the mechanisms through which risk is conferred. The recent limited efficacy of new disease modifying treatments (e.g., lecanemab) initiated after Alzheimer's disease (AD) is well established also have suggested the importance of starting treatments in people preclinical disease[3] if they could be identified with sufficient accuracy.

These questions cannot be answered fully with traditional observational studies of disease. Longitudinal studies of large, diverse populations are best suited to enabling the discovery science needed for advances in understanding risks and the preclinical progression of dementias. One of the first large, longitudinal population resources including data from participants of middle and older ages and open to researchers world-wide to enable this type of research is UK Biobank.

Here we describe this pioneering data resource, data and sample collection for which was started 20 years ago with joint funding from the UK's Medical Research Council and Wellcome Trust. UK Biobank includes participants who have gone on to develop the whole range of diseases of late life, including dementias.[1] The longitudinal design enables characterisation of the influences of risk factors and exposures preceding new diagnoses of a dementia. The comprehensive dataset, which includes family history, genetics, biomarker, imaging and clinical outcomes data, allows analyses of multiple risk or mechanistic determinants. The resource also can be used to define antecedent preclinical phenotypes. The population is now approximately 73 years old on average, a period in which risks of Alzheimer's disease roughly doubles in incidence every 5 years. Although UK Biobank was not designed specifically to study dementia, it is becoming a uniquely powerful resource for this.

Here, on the 20th anniversary since UK Biobank was started, we will describe the study design and principles of access. The size, breadth of phenotyping, linkage to clinical records that are continually updated, open availability of data and serial enhancement with support from the community of users highlight how it pioneered a new approach to large population epidemiological research. We will illustrate ways in which it has been used to study dementia to highlight major strengths of the resource. We also will reflect on limitations of the resource and describe some of the ways in which it is being improved for dementia studies. Finally, we will look ahead to ways in which the resource might be enhanced in ambitious ways that could accelerate translation of its findings to support lifelong brain health.

[1] https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=41270

UK Biobank: a unique resource for population-based studies of dementia

The size, breadth of data, long duration of follow-up and accessibility of the data for researchers make UK Biobank a powerful resource for advancing a broad range of studies related to the prevention, diagnosis, and treatment of common life-threatening and disabling conditions.[4] Potential participants were identified from UK National Health Service (NHS) patient registers between 2006 and 2010 and invitations to join UK Biobank were sent to almost 10 million adults who lived within travelling distance of one of 22 regional assessment centres. Sites for the assessment centres were chosen to ensure their accessibility to volunteers and the broadest possible demographic representation of the study population. Over 500,000 people aged 40–69 years were recruited across England, Scotland, and Wales.

This large sample size provides sufficient statistical power to assess relationships between multiple potential risk factors and a broad range of common health conditions of middle-to-old age[2]. While UK Biobank participants are generally healthier than the general population,[5] several studies suggest that exposure-disease relationships observed in the cohort are largely generalisable. Nonetheless, care must be taken in instances where volunteer bias could be a concern. Although the magnitude will vary between associations (e.g., those involving socio-behavioural traits may be more strongly affected[6]), participation bias in UK Biobank can influence effect sizes. Some studies have applied statistical approaches to adjust for volunteer biases in UK Biobank.[7,8] Care also should be taken when there is a potential for collider bias, when selection bias can substantially influence observed associations.[6]

At the time of recruitment, participants answered extensive questionnaires (including information on socioeconomic status, education, physical activity, diet, alcohol, smoking, sleep, and social engagement) underwent a broad range of physical assessments and contributed blood, urine and saliva samples. These samples have enabled cohort-wide genomic assays, including genotyping, whole exome and whole genome sequencing data and, for increasing proportions of the cohort, metabolomics (plasma, urine) and proteomics (plasma). Proteomic assays now are being performed for longitudinally collected blood samples to allow researchers to assess trajectories over time. Other assays, including those for serological markers of infection[9] and markers of immune response and transcriptomics of blood immune cells also are underway. When completed, following short embargo periods (whether the assays are funded by academic grants or industry) to allow the investigator consortia who funded the assays to explore them first, results from these will become available to all researchers who request them,

Of particular relevance for the study of dementia is the availability of a range of longitudinal cognitive function tests (including measures of reaction time, numeric and prospective memory, fluid intelligence) and the collection of imaging 3T MRI brain imaging (along with 3D carotid ultrasound and 1.5T cardiac and body imaging and dual-energy x-ray absorptiometry [DEXA]) on more than 100,000 participants with plans

[2] https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf, accessed 16 April 2025

to complete follow-up imaging assessments by 2028 on 60,000 of these participants over intervals 2-7 years following their first visits. The cognitive function tests have been validated psychometrically, can define population level relationships and have some predictive value for incident dementia.[10,11] However, they are not sensitive to individual variation or to specific functional system impairments (e.g., memory). Some discriminative power may be realised by repeat measures. Cognitive tests, along with lifestyle questionnaires, verbal interviews, blood sampling and a range of physical measures (including blood pressure and anthropometric measures) collected at baseline are being repeated at each imaging visit to assess changes over time in the imaging cohort. Environmental data, including exposure to pollutants, can be inferred by relating national meteorological databases to participants' postal codes. When linked to the family history and other questionnaire data and extensive clinical, genetic and other biomarker data included in UK Biobank (Fig.2), a broad range of risk factors for dementia can be discovered and potential causal mechanisms explored.

The prospective design of UK Biobank potentially reduces bias caused by "reverse causation" by enabling assessment of risk factors in many participants before dementias are diagnosed. UK Biobank currently has about 15 years of complete follow-up of participants' health via linkage to the NHS electronic healthcare records, including hospital inpatient records, death and cancer registry records. Primary care data are currently available for 45% of the cohort up to 2017. These data may be secured for the full cohort in time. Primary care data substantively augment that from electronic healthcare records with information concerning any family history of dementia (complementing the self-reported family histories), clinical diagnoses, clinical laboratory test results and medications prescribed. Cross referencing of hospital in-patient, death and GP records is needed for the most complete and accurate ascertainment of cases of dementia (Fig. 1). Consistent with expectations for earlier presentation of people with cognitive impairment to their GPs in the UK's health system, GP records include substantially more unique diagnoses of dementia than hospital records.[12] When combining these data sources, approximately 34,000 of the 500,000 participants are expected to develop dementia by 2030 based on current modelling by UK Biobank (Fig. 1).

In addition to UK Biobank's scale, depth of phenotyping and long follow-up period, what makes the resource so unique for dementia (and other) research is that the data are available via an open application process[3] to any *bona fide* researcher from anywhere in the world who is conducting health-related research that is in the public interest. Access to the data is provided to researchers from academic institutions and those from commercial organisations on the same basis.

Access fees (~£10,000 currently, depending on amount of data and numbers of sites in an application) are levied to recover costs incurred in providing access to the data. This upfront cost is perceived to be a

[3] https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access, accessed 16 April 2025

barrier to use of UK Biobank data for some. However, financial support or a full waiver of fees is available for early career researchers, students, and researchers from low- and middle-income countries[4].

When UK Biobank first made data available in 2012, researchers downloaded the data for their analyses. However, the subsequent development of secure, cloud-based analytical environments provided UK Biobank with an opportunity both to streamline access to the data (which, at over 30 petabytes, is now too large to disseminate) and to have more secure control over access to the participants' confidential data. A cloud-based UK Biobank Research Analysis Platform (RAP) was built. This enables researchers to access and analyse UK Biobank data *in situ* and ensures the highest levels of security and compliance with the UK General Data Protection Regulation[5]. The platform provides a suite of preloaded tools for data analysis such as R, Python, Jupyter Notebooks and other bioinformatics software, along with scalable computational resources for processing large datasets such as the whole genome sequencing data. Researchers can customise their compute environment, install analysis tools, and collaborate on the platform directly with other researchers approved for the same access application.

Preferential terms have been negotiated with the provider so that the RAP offers significant cost-savings for researchers relative to usual commercial cloud-based resources. Researchers also are only required to pay for storage of data they upload and derive or for their compute utilisation. To help researchers manage these costs, UK Biobank has received charitable funding to allow all researchers to receive 'research credits' for cloud use that can defray costs for at least some analyses with their UKB-RAP access approval. However, for university-based researchers who have access to substantial computing capacity at no cost through their institutions, RAP represents an additional cost that could not have been anticipated by either them or their funders.

A particular concern has been raised by researchers from the imaging community because of both the currently high estimated costs for large scale image analyses within the RAP and because of the relatively poor support for advanced imaging analyses within the RAP environment. UK Biobank and the provider are actively engaging with the imaging community to improve the latter. The current policy also is intended to be flexible. If the research involves analyses that simply cannot be performed on the UKB-RAP, researchers can apply for a project specific exemption for download of the data and analysis in a local environment. Finally, some researchers may be unfamiliar with programming and work within a cloud environment. There are a range of training resources for researchers on the web[6]. More are expected to be added in time (e.g., for image analysis).

UK Biobank already has had an impact on dementia research (Fig. 3, Table 1). In addition to studies of relative cognitive performance of heathy participants, a major focus of research with the still relatively

[4] https://www.ukbiobank.ac.uk/use-our-data/fees/financial-support

[5] https://www.gov.uk/data-protection, accessed 31 March 2025

[6] https://www.dnanexus.com/partnerships/ukb-rap-user-help-center/webinars

young UK Biobank cohort (the mean age at recruitment was 58 years old and the current median age of surviving participants is 73 years old [range 57-86 years old]) has been on understanding the prodromal stages of dementias and Parkinson's disease (PD), which is itself a risk factor for dementias (e.g., Parkinson's disease dementia [PDD], dementia with Lewy bodies [DLB] and AD). Related investigations to define genetic risks and the effects of to lifestyle and environmental exposures have been enabled by geographical and socio-economic diversity in the data.

New insights into determinants of dementia risk and disease mechanisms arising from UK Biobank

Genetic risks for dementia

Genetic data and its linkage to individual phenotypic and health data in UK Biobank have been a widely exploited enhancement to the resource. UK Biobank data have played a pivotal role in the identification of genetic risk loci for AD.

Early genome-wide association studies (GWAS) for AD included relatively small sample sizes (1000's of individuals) compared to the "mega-meta analyses" (hundreds of thousands of individuals) that are significantly enabled by UK Biobank data. Because there have historically been smaller numbers of participants with AD in the relatively young UK Biobank whose data could contribute to a genetic association study, the comprehensive family history data was leveraged for "diagnosis by proxy" in which those who self-reported one or both parents having a dementia diagnosis are considered as proxy cases. A number of studies have shown the utility of using an association study design based entirely on these proxy cases (in genome-wide associations by proxy or GWAX) by replicating association with loci identified by GWAS studies of clinically diagnosed cases.[13-15] Some concerns have been raised about potential biases this introduces.[16,17] Measurement errors in family health history surveys can lead to inclusion of both AD and non-AD dementia cases and thereby attenuate genuine genetic associations for AD. Combining GWAS and GWAX without accounting for heterogeneity among the associations can lead to substantial downward bias in heritability estimation[18]. However, there are methods to mitigate these issues such as exclusion of individuals who have not reported or do not know about their parents' health from analyses, weighting proxy case status by the number of parents with a dementia diagnosis[19] and controlling for parental age. The most recent AD and related dementias (ADD) combined GWAS/GWAX included data from 111,326 clinically diagnosed or "proxy" cases of AD (46,828 of these were "proxy" cases from amongst the UK Biobank participants) and identified 75 risk loci in total, of which 42 had not been reported by previous GWAS studies.[20]

These data have substantively informed current hypotheses regarding mechanisms for genesis of the disease. A key observation has been the enrichment of genetic risk loci in specific biological pathways, such as for amyloid, tau, lipid processing, endocytosis and innate immune responses, highlighting their potential roles in the genesis of AD.[15,20,21] [15,20,21] Leveraging the more recent availability of whole exome sequencing data in UK Biobank has enabled large-scale analyses of rare variation resulting in the identification of additional novel risk loci for AD.[22,23] Data also have been used to develop polygenic risk

scores (PRS) predictive of future dementia risk (Box 1)[20,24] and in Mendelian randomisation (MR) analyses, whereby genetic variants are used as instruments to assess causality of an observed association

Digital biomarkers for disease risk prediction

Passively acquired digital measures of activity are providing powerful measures of variety of phenotypic feature that are sensitive neurodegenerative disease or risk. Wearable sensors, such as wrist-worn activity trackers, allow for passive data collection and therefore do not rely on participant recall, and they are minimally invasive and easy to use.

Between 2013 and 2015, over 100,000 UK Biobank participants wore Axivity Ax3 wrist accelerometer devices for one week to objectively measure physical activity throughout the day. The most fully developed work to date has explored Parkinson's disease (PD); researchers have used these data to discover measures of dementia risk and brain health based on features that can be derived from these data such as sedentary time,[25-27] average acceleration,[27-29] step count[30] and circadian rest-activity rhythms.[31-34] Research using UK Biobank data has highlighted the potential health benefits associated with moderate-to-vigorous physical activity (MVPA) in mid-to-older age, which included a 45% reduced risk of PD and reductions in risk for future dementia. [35] [29] A more recent nested cohort study design using UK Biobank (Fig. 4A) has discovered specific accelerometry features sensitive to early or prodromal PD: lower average acceleration with movement was found to be sensitive to risk of future disease in participants up to seven years before a clinical diagnosis of PD was made (Fig. 4B).[35] The wide range of data in UK Biobank allowed investigators also to show that accelerometry data may outperform other predictive biomarkers considered independently (genetics, lifestyle factors, and blood biomarkers) in early prediction of a diagnosis of PD (Fig. 4C).

PD is of interest in the context of this review because PD shares risk factors and neurodegenerative mechanisms with other disorders in the ADRD spectrum and is a risk factor for AD. [32,36] In addition to reduced daytime movement acceleration, sleep-related features such as a shortened consecutive sleep duration distinguish preclinical PD from other neurodegenerative conditions (Fig. 4).[35] Moreover, accelerometry-derived features may add considerable additional information to that available from genetic, lifestyle, and blood biomarkers (Fig. 4C). Different features of the accelerometry time courses may be sensitive to different diseases, e.g., analytical metrics that can be derived from the same data were shown able to predict the future onset of AD in the UK Biobank population.[32]

However, these studies also highlight some general limitations of the UK Biobank data for the study of prodromal disease. To date, as primary care data are linked only for a minority of participants, researchers must place greater reliance on hospital inpatient data, which are likely to record diagnoses later, and the relatively sparse medication records (e.g., only ~5.4% of PD cases documented use of symptomatic treatments).[35] Consequent apparent delays in recorded diagnoses (~2.3 years for PD) can inflate predictive models. Improving data accuracy and more comprehensive access to primary care records will be critical

for enabling future studies. There also are limitations specific to the accelerometry data. For example, a single one-week accelerometry data collection period does not allow for the assessment of long-term progression. There is a strong case for longer-term or repeated movement data capture from UK Biobank participants in the future. The availability of clinical, genetic, exposure and a growing amount of multi-variate plasma biomarker data offers the opportunity for future exploration of whether more accurate or specific early diagnoses can be made by optimising the integration of multiple risk factors and potential indices of early disease expression.

UK Biobank brain imaging for pathology-related gene discovery and pre-clinical disease characterisation

The UK Biobank multi-modal imaging enhancement (2014- date) is acquiring brain and body MRI imaging, 3D carotid ultrasound and DEXA from over 100,000 participants and approximately two-thirds are being invited back for a second scan. This will allow measures of longitudinal brain (and body) changes over interscan intervals ranging from 2-7 years.[37] The different types of brain MRI that are being used provide measures of brain structure, function, connectivity and microstructure.[38] Although conducted across four imaging centres, a strong focus on harmonisation of the scanners and imaging protocols and close monitoring of image quality have supported generation of an exceptionally high quality dataset.[39] The unprocessed image data and over 4000 imaging-derived phenotypes (IDPs), each providing a measure based on a different brain structural or functional feature, are available in the database for researchers to use.

The cost, logistical complexities and concern for the impact of adding a radioactive biomarker test on subject recruitment have so far precluded the inclusion of amyloid or tau PET scanning. However, MRI provides other markers related to relevant disease processes.[40,41] The new longitudinal imaging data will be of particular interest for dementia research because it allows more sensitive, individually normalised measures of progressive regional brain atrophy, which shows patterns distinct for different types of dementia (e.g., early relative hippocampal atrophy for AD vs. early relative fronto-temporal cortical atrophy for frontotemporal dementia), or changes in white matter T_2 -hyperintense lesion counts in cerebral small vessel disease (cSVD, Box 2).[40] Additional MRI-derived measures such as functional connectivity (an index of brain activity) or quantitative magnetic susceptibility (QSM, which is sensitive to differences in tissue iron and myelin) are examples of complementary markers sensitive to phenotypic variation in the population and to early dementia or other neurodegenerative disease.[42-44]

Because of the large dataset and consistently acquired imaging data, even weaker associations of imaging derived measures and phenotype or genetics can be discovered and researchers have been able to discover evidence for previously unrecognised aspects of preclinical stages of dementia.[45 46] For example, an exploration of the age-dependence of associations between the genetic risk score for AD and localised brain atrophy consistent with early neurodegeneration discovered trends beginning as early as middle age (45-53 years old) for “trans-modal” limbic, subcortical, hippocampal and orbitofrontal grey matter

differences that could be related to genetic variation in the *MAPT* gene cluster (Fig. 5).[47] This interpretation has been strengthened by independent evidence for causal relationships between structural brain imaging signatures of AD or frontotemporal dementia and clinical disease based on an independent two-sample MR analysis.[48] Other efforts are employing new imaging derived measures of brain differences in disease to enhance the discovery of novel genetic associations from which new mechanistic hypotheses can be derived, such as that higher iron deposition in subcortical brain regions contributes to neurodegeneration in Parkinson's disease and non-Alzheimer's dementia.[49,50] As an intermediate phenotype, brain imaging can discover influences of potentially modifiable risk factors and identify brain regions that mediate relationships between genetic variation and dementia.[51] An illustrative study used lifestyle measures to calculate a surrogate measure of cognitive reserve and then to demonstrate that cognitive reserve modulates relationships between cardiometabolic disease and dementia and brain volume loss.[52] Other studies have provided evidence for interactions between dementia risk factors (e, g., alcohol use, diabetes and air pollution) and ranked their relative impacts on brain structure and dementia.[51,53,54] Together, these highlight the value of imaging measures as for mechanistic hypothesis generation concerning genetic and other risk factors.

Emerging proteomic and lipidomic biomarkers of disease and risk

To date, the UK Biobank Pharma Proteomics Project consortium (UKB-PPP) has supported measurement of 2923 blood plasma proteins in over 53,000 participants. Currently, UK Biobank is the world's largest population dataset for exploring relations between plasma protein levels, health and disease.[155] A number of studies already have tested models for prediction of future dementia and other diseases based on multivariate plasma protein profiles measured from samples collected during baseline assessments of the UK Biobank participants.[56-58] Well established markers of brain neuronal damage (neurofilament light chain, Nfl) and reactive astrocytes reactivity (glial fibrillary acidic protein, GFAP) were found to change more than 10 years prior to dementia diagnosis.[58] The accuracy of proteomics measures to predict dementia independently was similar to combinations of genetic, health and lifestyle measures.[56] MR has been used to identify CEND1 and SYT1 as having potentially causal roles in AD.[59] Separate analyses of body and organ proteomic “aging clocks” demonstrated that proteomically “older” individuals have a higher likelihood of future dementia diagnosis and separately mapped differential contributions of genetics and exposures to organ-specific aging.[60,61] An analysis of the association between poor cardiovascular health and incident dementia identified growth/differentiation factor 15 (GDF15), a plasma peptide that is associated with aging, as a mediator of this association.[62] Much more can be expected in the near future with the 2025 announcement that UKB-PPP will further enhance the resource to include assay data on 5,400 proteins using the most recent version of Creative Proteomics antibody-based Olink technology, Olink Explore HT[7]. These assays will be conducted on baseline samples from all of the UK Biobank participants and on the 100,000 repeat samples taken up to 15 years later from the subset of volunteers to

[7] https://olinkpanel.creative-proteomics.com/panel/olink-explore-ht-panel.html?gad_source=1&gad_campaignid=22941706220&gbraid=0AAAADopmboCPRuLld8oOwAupFdbTExZZ&gclid=CjwKCAiAlrXJBhBAEiwA-5pgwjx-_vLI3y4UB4-1YJJXMzczA3GbirBodRIBBoC7B8V135Kqw0ubtRoCjX0QAvD_BwE

the imaging enhancement.[8] In accordance with standard access sample procedures for UK Biobank, the fourteen biopharmaceutical companies funding the project will have a 9-month exclusivity period for use of the assay data before it is made available to all approved researchers.

Plasma nuclear magnetic resonance (NMR) measures of 250 lipoprotein lipids, fatty acids, and small molecules such as amino acids, ketones, and glycolysis metabolites were made available for 300,000 UK Biobank participants in 2023, with data on all 500,000 participants made available at the end of 2025.[63] These measures have been used to characterise metabolomic profiles for the different dementias.[64,65] Plasma metabolites profiles are sensitive to gene variation associated with risk of AD.[66] Subsets of the metabolites (including, e.g., lipoprotein lipid concentrations, linoleic acid, sphingomyelin, glucose, and branched-chain amino acids) were found to be predictive of future dementia diagnoses. Similar associations were identified in analyses of serum lipids across 470,294 participants.[67]

Modifiable risk factors for dementia

There is major public health interest in identifying risk factors that, if reduced, could lower the future incidence of dementia. The 2020 and 2024 *Lancet Commission* reports on dementia identified 14 modifiable lifestyle and environmental risk factors for later-life dementia.[68,69] The initial report estimated that around 45% of cases of dementia are potentially preventable. The size, depth of characterisation and heterogeneity of exposures of participants in UK Biobank already have made it a singularly important resource for characterising environmental exposures and other potential modifiable risk factors for dementia (Table 1).[27,70] The potential clinically meaningful impact of their modification for prevention of dementia also has been highlighted in a recent comprehensive analysis of their joint associations with dementia. Even so, associations discovered need to be interpreted with consideration of the role of reverse causation bias even with the longitudinal observation possible using UK Biobank: observed associations may be indicative of changes in behaviour during preclinical and prodromal stages of dementia, rather than causal associations for primary prevention. [71,72] Some of this concern can be mitigated by using MR to test for evidence of causal relationships between dementia and genetic proxies of modifiable risk factors[50], although limitations imposed by the specific assumptions of MR need to be considered in each case.[73]

Discovering new modifiable risk factors

The *Lancet Commission* considered that evidence for sleep disturbance and diet was insufficient to include them in the list of modifiable risk factors for dementia despite results suggesting associations.[69] These have been active recent areas for further investigation using UK Biobank. For example, self-reported long sleep duration (>9 hours/night) was associated with a higher risk of AD independent of the AD genetic risk.[74] Objective measures of sleep from accelerometry have shown associations of deterioration in measures of

[8] www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/launch-of-world-s-most-significant-protein-study-set-to-usher-in-new-understanding-for-medicine

sleep with Parkinson's disease and other dementias[35] and another study suggested a U-shaped association between sleep regularity and incident dementia.[33] MR studies have provided additional evidence that both extremes of short and long sleep duration across the participants reduce cognitive performance.[75]

Computerised 24-hour diet recall questionnaires in UK Biobank (Oxford WebQ), also have been used to investigate the effect of diet on risks of dementia. For example, habitual consumption of the Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) diet,[76] which emphasises consumption of vegetables, fruits, whole grains, and healthy fats, was associated with a reduced risk of Alzheimer's disease.[77] Consistent with this, a study of a subset of the cohort showed that higher adherence to the related Mediterranean diet was associated with decreased dementia risk.[78] However, MIND diet adherence in a larger group found evidence for a lower dementia risk only in women.[79] This illustrates the opportunities provided by UK Biobank for researchers to re-assess analyses performed and published by others using the same data. Over time, this will contribute to much greater confidence in consensus results generated based on different, independent analyses.

Differences in dietary associations for different types of dementia have been found. For example, fish oil supplement use was associated with reduced risk of all-cause dementia and cSVD (but not AD), although the apparent benefit was attenuated with increasing *APOE* ε4 dosage.[80] Complex relationships between dietary factors and dementia demand close scrutiny in the future because of their potential dependence on the model used for analysis. For example, J-shaped or U-shaped associations were found between risks of all-cause dementia and the consumption of coffee or tea, respectively in hypertensive participants.[81] An important caveat to all of these studies is that they are based on self-reported diet diaries taken at one point in time and, as observational studies, cannot fully correct for confounds. As a potential example of the latter for the case of coffee consumption, despite the evidence for association, Mendelian randomisation across UK Biobank and other European cohorts failed to provide evidence for any potentially causal relationship between coffee consumption and late life cognition. [82]

The latter results highlight the fundamental strength conferred by the open access policy of UK Biobank: multiple research groups can rapidly test the reproducibility of new studies using the same data. This encourages "self-correction" by the scientific community. For example, a recent study linking hearing aid use to dementia risk was retracted after an independent group attempted unsuccessfully to replicate the findings, which were discovered later to be confounded by a computational error.[83].

UK Biobank is helping to extend the range of modifiable risk factors identified. A particularly topical area of exploration has been for links between infections and later dementia diagnoses.[84] Hospital-treated infections were associated with raised rates of both AD and all-cause dementia.[85] A comprehensive study of associations between neurodegenerative diseases and virus exposures across UK Biobank and Finland's FinGenn biobank found 22 replicated virus-disease associations with AD, PD, multiple sclerosis, amyotrophic lateral sclerosis or cSVD diagnoses with elevated risks persisting for up to 15 years following

infection.[86] UK Biobank's Infectious Disease Pilot Study, which provides serology for 20 pathogens in 10,000 participants, was used to discover increased risk for dementia in participants seropositive for HSV1 and other viruses of the Herpesviridae family.[9] Records of influenza vaccination in the Biobank were associated with a reduced risk for incident dementia.[87] Enabled by a series of initiatives allowing UK Biobank to acquire additional data to study impacts of SARS-CoV-2 on its volunteers early in the 2019-2023 pandemic, evidence for mechanistic relationships between effects of infection on the brain leading to acceleration of pathology related to AD was reported recently (Box 3). Further enhancements to UK Biobank using the extensive biosamples collected, such as the new plasma proteomic assays and an expanded range of markers of exposures to infectious diseases, are likely to add further to understanding of the associations discovered and the underlying mechanisms responsible.

Future opportunities to enhance UK Biobank for future dementia studies

Major strengths of UK Biobank for discovery of risk factors for dementia are the size, longitudinal design and the breadth and depth of data acquired on participants. The scientific value of UK Biobank data for dementia research also continues to grow with the increasing number of incident cases as the population ages. Moreover, stored biosamples from participants, including repeated samples from large proportions of the cohort, are available for additional assays. For example, as methylation patterns on the genome are responsive to exposures, currently planned epigenetic assays (using long-read sequencing) on 50,000 participants should enable better estimations of lifetime environmental exposures and their impacts on future clinical outcomes. Additionally, as DNA methylation increases with biological age, they can provide new insights into ageing-related physiological processes. In conjunction with what already is available, these new data will offer exciting opportunities to gain deeper insights into the molecular mechanisms underlying different causes of dementia and enable identification of biomarkers for earlier diagnoses or risk stratification and discover novel candidate therapeutic targets.[57-59,88]

UK Biobank is at an inflection point in its utility for dementia studies. Numbers of incident cases are rising sharply as the participants age (Fig.1), increasing the statistical power to explore associations with lifestyle exposures. However, fundamental limitations for dementia research that need to be better addressed if the resource is to be made more useful are the current relative insensitivity of cognitive data to early disease, the completeness of ascertainment of participants with dementia and the accuracy of diagnoses for cases reported.

There are promising developments to address these issues. With implementation of the RAP, data security requirements have been met that should allow access to primary care data on all UK Biobank participants. The inclusion of primary care data for all participants would double the number of people identified with dementia and will allow identification of dementia at earlier stages of progression (Fig. 1).[12,89]. Online cognitive testing batteries more sensitive to age-related cognitive decline and impairment on an individual level now are available[90] and plans for their addition online delivery to participants are well advanced.

But more will need to be done. For example, even with this further case ascertainment from medical records, about half of dementia cases are likely to be unspecified, with additional misdiagnosis and underdiagnosis. To better address this and other current limitations of UK Biobank for dementia studies, a new initiative, the UK Biobank Brain Health Study, is being developed to enhance the phenotyping and specificity of diagnoses of participants with established disease The assessment ultimately will involve recall of thousands of UK Biobank participants per year who have been identified via their medical records or self-report as having diagnosis of new onset cognitive impairment or dementia for additional cognitive assessments (including the MoCA as a measure of global cognition), detailed sensory and motor function examinations, additional brain MRI scanning and blood sample collection (to measure current levels of biomarkers of particular relevance to neurodegeneration, such as brain-derived tau protein isoforms, amyloid betas, neurofilaments) during a specialised assessment centre visit, followed by additional data collection remotely. With success, there is the potential for extension of the procedures conducted during assessment visits to include cerebrospinal fluid sampling for a subset of the cohort, data from which could be used for interpretation of plasma biomarker measures more widely in UK Biobank. More ambitiously, amyloid PET scans might be able to be conducted on a sub-population of the Brain Health study participants in time.[91] This would enable, for example, diagnoses of limbic-predominant age-related TDP-43 encephalopathy (LATE), a recently defined and common dementia that mimics the clinical symptoms of AD.[92]

Use of mobile app technologies present another promising opportunity to collect detailed information on physical activity, sleep patterns, speech and spoken language, light exposure and geo-location for large proportions of the participants. This approach may also be particularly well suited for people with dementia, as many already use mobile devices in their daily lives. Patterns learned from wearables are expected to improve pre-morbid differentiation between people at risk or with AD, dementia with Lewy bodies, and PD dementia to refine future genetic and risk factor association studies.[93,94] Expanding linkages to other datasets, such as those for including detailed outdoor air quality data, will make the resource even more valuable and relevant to emerging areas of interest.[95] Finally, developing plans to invite all surviving UK Biobank participants for another full in person assessment would allow all of the data collected at baseline to be repeated with some additional data related to frailty and ageing (e.g., more comprehensive cognitive and motor assessments), vision (OCT and eye tracking) and hearing. This would be singularly valuable for enabling research into the determinants of prodromal neurodegenerative disease and the development of early disease biomarkers as the population is aging and the incidence of dementia is rising. As long as the extraordinarily high commitment shown by UK Biobank participants is maintained, the primary practical limitations to these additional enhancements are simply their costs when scaled to the size of the cohort and the logistical challenges of delivering them efficiently and in a way that maintains a positive experience for volunteers.

Conclusion

UK Biobank initially was highly controversial.[96] Prominent geneticists argued that more targeted studies could deliver genetic associations more quickly, the use of research data for research that would be defined in the future was argued by some critics to be inconsistent with principle informed consent and the plans to treat data access for companies and academic research scientists on equal terms was highly contentious. Only a strong lead from the Wellcome Trust and the UKRI MRC allowed its successful initiation.

Since then, the value of genetic investigations of a deeply, precisely phenotyped population has been shown repeatedly, despite its heterogeneity. The model of broad consent with continued oversight by an independent ethics and governance committee and protected by the right to for participants withdraw from their data from the resource at any time has been accepted by participants, researchers and the wider public. The open model allowing *bona fide* researchers from academia or industry access to the data on equal terms - as long as they contribute their derived measures back to UK Biobank to be used by others - has been shown to be hugely successful. Repeated examples of industry support for major enhancements to the resource (e.g., for genetics, proteomics and imaging) that benefit all users and the more than 18,000 peer reviewed publications citing the resource as of mid-2025 are testimonies to this success.

As the population ages, the value of the resource for dementia research is rising with numbers of incident cases of dementia. The size, depth of characterisation of the participants, and availability wide clinical outcomes data have enabled UK Biobank uniquely. This population-based study provides a powerful adjunct to disease-cohort natural history studies. Moreover, it directly addresses the critical needs to understand the genesis of dementias and to discover effective preventative strategies. With rapidly rising numbers of older people at risk, particularly in the developed world, public health systems will not be able to cope with the costs of care and treatment of established disease, even if effective approaches for the latter are found.

As UK Biobank enters its third decade, we encourage researchers to contribute ideas and lead in raising new funding to enable UK Biobank to make step changes in its capabilities for dementia research. Given the breadth of data already available on cohort participants, the infrastructure in place for management of the study and the remarkable commitment of the participant community, further investment in UK Biobank will deliver high value. There is a growing opportunity to exploit this resource for dementia research.

Tables

Table 1. Examples of how data from UK Biobank has contributed to better defining modifiable risk factors for dementia identified by the Lancet Commission. [68,69]

Risk Factors	Illustrative examples of related studies
Depression	Supporting the hypothesis of modifiability of risk, both an increased risk of developing dementia for UK Biobank participants who reported depression and a relatively reduced risk for those with treated depression was found.[97]
Hearing and vision loss	Amongst dementia-free participants aged ≥60 years, there was increased risk of dementia among participants with objectively measured poor speech-in-noise hearing.[98] There also was both an increased risk of dementia in those with cataracts and a lower dementia risk in others treated surgically.[80]
High LDL cholesterol	An association of lower testosterone levels with increased AD risk was shown to be mediated through increases in low-density lipoprotein cholesterol (LDL) and obesity,[99] a finding strengthened by follow up two-sample MR studies that show a 2-3 fold increased risk of developing AD per standard deviation increases in blood levels of total cholesterol and of LDL-cholesterol.[65]
High blood pressure	UK Biobank studies have highlighted heterogeneity in the influence of hypertension on different dementias. Earlier onset of hypertension was associated with a higher incidence of all cause late-life dementia and MR suggested a causal relationship. However, an independent study showed that genetic proxies for hypertension were associated with a reduced risk of AD.[100,101] A recent preprint has used MR to confirm an association of genetically proxied systolic blood pressure with an increased risk of all types of dementia and vascular dementia, but not AD.[102]
Obesity	While an initial study concluded that higher body mass index (BMI) at the baseline visit was associated with a lower risk of dementia after 8 years of follow up,[4] a subsequent study of a subset of the cohort with a mean of 8.7 years of follow up found evidence for a U-shaped association between BMI and dementia: more upper limb fat and less fat-free mass was associated with a higher risk of dementia.[103] Two subsequent studies focusing on participants with a low muscle mass-high body fat phenotype (sarcopenic or dynapenic obesity) has extended this observation, reporting that obesity with low muscle mass is associated with increased dementia risk particularly for women, participants < 65 years old and those who do not carry the *ApoE4* risk allele.
Smoking	Smoking was associated with an increased risk of dementia, with the highest age-specific hazard ratio (HR) occurring in the youngest age group (40 - 49 years old at

	baseline) in UK Biobank.[104] Early-life tobacco exposure, including maternal smoking during pregnancy, was also associated with an accelerated onset of dementia.[105] Effects of a smoking history on dementia risk were modified by the apolipoprotein E (*APOE*) genotype in older participants (60-73 years old at baseline).[106]
Social isolation	Social isolation amongst UK Biobank participants, but not loneliness, was associated with an increased risk of dementia. Socially isolated individuals also had lower grey matter volumes on brain MR.[107]
Diabetes	Compared to people without diabetes, people with type 2 diabetes in UK Biobank had an increased risk of all-cause dementia. Associations were strongest for women and amongst those with a younger age at diabetes onset.[108,109] There also was a trend for people whose diabetes was diagnosed before age 55 to have a higher risk of developing cSVD. However, despite robust associations of diabetes with AD, genetic risk determinants for diabetes do not contribute significantly to AD risk.[110] Together, these results suggest that diabetes potentiates other more directly acting risk factors potentially related to recognised complications of diabetes that themselves are responsible for doubling the risks of all-cause dementia and AD.
Lower levels of education	Several studies using UK Biobank data have reinforced a positive association of more education and a higher socio-economic status (SES) with better cognitive health.[52,109,111-113] Higher cognitive reserve (as defined by greater education and participation in cognitively engaging occupations and activities at the time of assessment) was associated with a reduced dementia risk and slightly delayed onset of dementia on average.[113]
Physical inactivity	UK Biobank participants with the highest compared to the lowest fifth of reported leisure-time physical activity were less likely to develop dementia, while those who were most sedentary based on self-reported data had a higher risk. This was confirmed by objectively measured accelerometery: higher daily MVPA of just 14.5 minutes was associated with lower risks of AD comparable to two years of aging.[26,30,114,115]
Excessive alcohol consumption	An early study using MR suggested a causal (linear) relationship of alcohol intake with dementia risk.[116] Concordantly, there was a continuous relationship between greater alcohol consumption and lower brain volumes across a full range of alcohol consumption, suggesting that any alcohol intake reduces brain health.[117,118] However, another analysis provided evidence for heterogeneity in responses across the population: "non-moderate" patterns of alcohol intake by participants in UK Biobank of either kind– either zero or >2 (men) or >1 (women) drinks per day – were associated with increased risks of dementia but risks varied with age, being higher in younger participants (40 to 49 years at baseline).[104] Together, these provide an illustration of the value of open data for multiple, independent analyses to better define levels of confidence in conclusions than is possible from single studies.

Air pollution	Exposure to multiple air pollutants ($PM_{2.5}$, PM_{10}, NO_2, and NO_x) was associated with an increased risk of dementia, especially among participants with high genetic susceptibility.[119] Risks for different dementias may vary; elevations in $PM_{2.5}$ are associated with greater risks of all-cause dementia and AD, while NO_2 is associated with greater risk of any incident dementia, AD and cSVD.[51,95] A healthy lifestyle may favourably modify this.[120,121]

622

623

Text boxes

Definitions

AD	Alzheimer's disease
ADRD	Alzheimer's disease and related disorders, a general term that includes Alzheimer's disease and other disorders such as frontotemporal degeneration (FTD), Lewy body dementia (LBD), and cerebral small vessel disease (cSVD) that share late-life impairment of cognitive performance across multiple dimensions (e.g., memory, reasoning, and language) with neurodegenerative pathology. Parkinson's disease dementia (PDD) often also is included.
BMI	The body mass index, calculated by dividing an individual's weight in kilograms by their height in meters squared (kg/m^2), provides a simple index of how body mass related to body composition.
cSVD	Cerebral small vessel disease is a heterogeneous group of disorders associated with narrowing or blockage of brain microvasculature, classically related to stroke, gait disorders and cognitive impairment.
DEXA	Dual-energy x-ray absorptiometry, a clinical imaging method that uses two very low energy X-ray beams of different wavelengths to measure bone density and body relative lean and fat mass.
GP	A primary care doctor in the UK. Typically, the first regular point of contact for people in the UK with general health problems needing medical attention.
GWAS	A Genome-Wide Association Study, in which DNA single nucleotide polymorphisms (SNPs) of people with a diagnosis of a disease of interest are contrasted with those from people who do not have the diagnosis. Differences in associations define those SNPs related to the expression of the disease or phenotype of interest. SNPs can be related to the structure or expression of genes.
GWAX	A Genome-Wide Association Study by proxy is a variant of the GWAS that uses a proxy (e.g., family history of a disease) instead of an individual's expression of that disease or phenotype to identify associated SNPs. In the case of a late life disease like AD, larger numbers or younger populations can be used to discover associated SNPs as this approach does not rely on clinical disease expression.
HR	Hazard ratio, a ratio expressing the frequency of an outcome (e.g., disease) typically in a disease population relative to a healthy or other reference group.
IDP	An Image Derived Phenotype is a quantitative feature derived from imaging that describes some aspect of individual phenotype.
MR	Mendelian randomisation is an analytical approach to testing for a potential causal relationship between a risk factor or exposure and a health (or other) outcome of interest. It relies on the prior independent identification of SNPs associated with both factors. The method relies on the random inheritance of genes and is useful when random allocation of subjects to comparator groups for testing comparator groups is not practical.
MVPA	Moderate-to-vigorous physical activity

PD	Parkinson's disease
$PM_{2.5}$	Heterogeneous microparticulates, typically in air, with a diameter of 2.5 μm or less. Small particulates in this range are associated with adverse effects on human health.
PRS	The polygenic risk score estimates the magnitude of an individual's genetic predisposition to a complex disease. It represents a simple sum of the genetic risks conferred by each of the SNPs carried by that individual which are associated with a disease.
SNP	A Single Nucleotide Polymorphism is a single nucleotide difference in DNA sequence that is found in at least 1% of the population.
QSM	Quantitative Susceptibility Mapping refers to images generated using MRI contrast based on phase of gradient-echo MRI data. The maps of local magnetic susceptibility, a measure of the magnetic field, are highly sensitive to the distribution of cell iron and myelin.

Box 1. Polygenic risk stratification for dementia

Polygenic risk scores (PRS) provide a quantitative measure of an individual's genetic risk for a given disease. Risk prediction may be particularly useful clinically for late-onset disorders such as dementia as individuals at high risk of developing dementia can be identified at an earlier stage, potentially enabling presymptomatic intervention.[20,122] They also can contribute to the identification of more homogeneous subgroups of individuals to enrich clinical trials or for personalisation of clinical management. An individual's PRS is calculated by summing the alleles of interest weighted by their effect sizes for association with the disease. In a small study including neuropathologically confirmed cases of AD,[20] assessing the PRS together with separate modelling for *APOE* contributions was reported to have a prediction accuracy approaching 80% in the general population and up to 95% in individuals at the extremes of high or low risks for dementia.[123] However, clinical prediction accuracy may be lower in practice not only because of the small size of group studied but also because effect sizes for AD are typically smaller in clinically diagnosed groups relative to those that are neuropathologically confirmed[124]. The specific risk contribution is age-dependent.[123] Pathway specific PRS estimates which reflect the genetic risk in a particular biological pathway also can be derived and used to link specific biological differences to relative genetic risks.[125,126] The predictive utility of PRS for AD can be improved by the addition of other biomarker data.[127] Correlation of brain imaging data with AD PRS has defined structural variation with AD PRS that putatively reflects preclinical brain pathology in people who are without clinically expressed symptoms.[125] Calculating PRS in large-scale, deeply phenotyped datasets that include a range of additional measures sensitive to disease expression (such as UK Biobank) provides a way of evaluating the potential future utility of PRS in the clinic.

Box 2. *Brain imaging-based genetic association studies and cSVD*

The impact of UK Biobank imaging data has been particularly great for understanding dementia related to cerebral small vessel disease (cSVD).[128] Covert cSVD, which is diagnosed based on predominantly white matter brain MRI lesions in the absence of clinical stroke (based on white matter hyperintensities, cerebral microbleeds and enlarged perivascular spaces), is highly prevalent in the general population, particularly with increasing age. UK Biobank has played a major role in unraveling genetic correlates of different cSVD phenotypes. Because of its population-based design, this has been true especially for covert cSVD. Large collaborative genome-wide association studies (GWAS) integrating data from UK Biobank with other, smaller datasets (e.g., from the Cohorts for Heart and Aging Research in Genomic Epidemiology[129] and the International Stroke Genetics Consortium[130]) have discovered over 70 independent cSVD risk loci, mostly for MRI-cSVD[83,131-134] but also for cSVD-related stroke.[135] *In silico* functional explorations of these using UK Biobank data point to a major role for blood pressure-related pathways[133] and demonstrated that genetic determinants of cSVD are shared with those for aortic distensibility.[136] Pathways distinct from those for known vascular risk factors have been revealed, including those for the cerebrovascular matrisome, membrane transport, vascular development, myelination, and the blood-brain barrier.[131] In addition, the diversity of MRI sequences from which data is available in UK Biobank enables exploration of the genetic

architecture of novel cSVD imaging proxies, such as measures of white matter microstructure assessed using MRI diffusion imaging or features believed to reflect function of the glymphatic system (e.g., the perivascular space burden or water diffusion in the perivascular space).[46,132,134,137-140] Beyond GWAS, the availability of whole-exome sequencing data on all participants in UK Biobank was instrumental in revealing a genetic continuum between monogenic and multifactorial cSVD by showing that several genes (*HTRA1, COL4A1/2*) harbor both rare mutations and common variants associated with the disease and that disease-causing mutations are less rare in the population than previously thought.[134,135,141,142] Screening for human knock-outs, exploring the functional impact of mutations affecting specific protein domains and assessing their phenome-wide correlates in the UK Biobank provided insights into the biological mechanisms underlying cSVD and, in some cases, potential effects of their therapeutic modulation. [83,135,141,143] More recently, these imaging genetics association studies have begun to leverage other resources in UK Biobank (e.g., proteomics, metabolomics) that are enabling identification of fluid biomarkers of disease risk[144,145,146,147] and promise to accelerate 'omics-driven drug discovery.[83,135,148]

Box 3. SARS-CoV-2 and dementia

UK Biobank undertook special initiatives during the 2019-2023 COVID-19 pandemic that provided new resources for studying the potential risks of a coronavirus infection for dementia[9]. Lateral-flow tests and SARS-CoV-2 serology was acquired for large numbers of UK Biobank participants and access was provided to their national SARS-CoV-2 test results and vaccination records. These data have contributed to demonstrating that dementia was a major risk factor for both SARS-CoV-2 infection and subsequent hospitalisation and mortality. [149-151] *APOE4* genotypes were found to confer additional risks of severe clinical outcomes with infection.[152] Over 1000 participants in UK Biobank who had undergone imaging assessments soon before the start of the pandemic and who had evidence for SARS-CoV-2 infection by late 2020 or were demographically matched controls with no evidence of infection were invited for repeat clinical and imaging assessments in early 2021. These data provide a unique pre- and post-infection case-control matched dataset that allows acute to sub-acute effects of infection to be studied. Participants who had experienced infections, despite most of them having had only mild to moderate symptoms, showed a relative cognitive impairment after infection. This was associated with regional temporal lobe brain atrophy[153] and interval changes plasma proteomic biomarkers in a direction consistent with greater brain β-amyloid pathology. These data suggest that SARS-CoV-2 infection (and potentially other viral infections associated with a substantial systemic inflammatory response) may potentiate early AD-related brain pathology - despite a lack of evidence for neurotropism.[144]

[9] www.ukbiobank.ac.uk/learn-more-about-uk-biobank/covid-19-hub

Figures

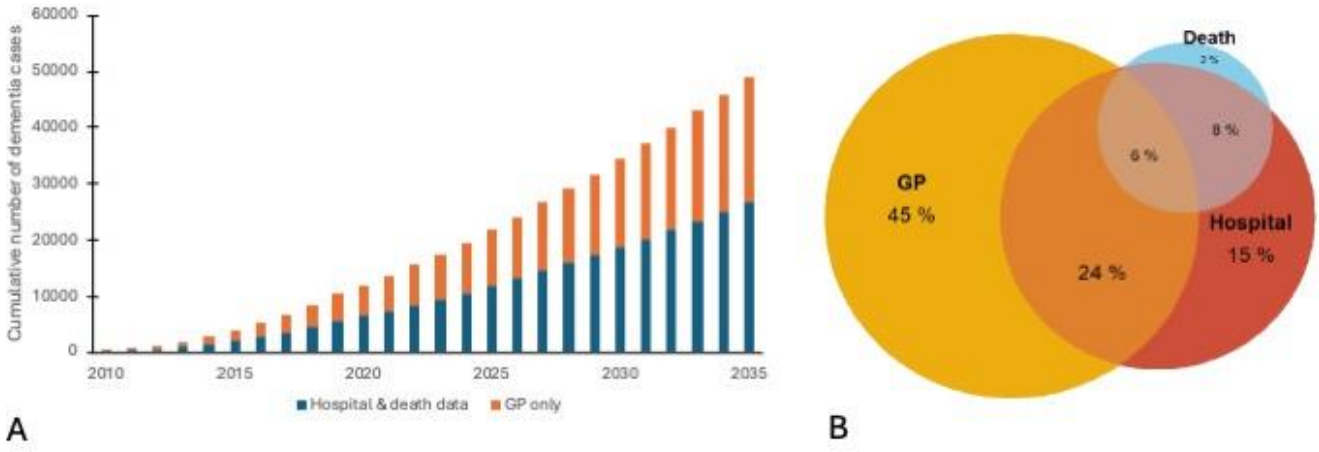

Fig. 1. (A) Estimated growth in cases of dementia among UK Biobank participants over time. By the end of 2020, there were 6,458 cases of dementia identified from hospital in-patient admissions and death registry data (blue bars). Inclusion of primary care data nearly doubled (as shown by orange bars) the number of cases identified (12,000 total). Conservative estimate projections suggest that there will be a steady rise in the number of UK Biobank participants with diagnoses of dementia over the next five years with 34,000 incident cases expected by 2030. (**B) Venn diagram illustrating relative contributions of different data sources (GP records, hospital episode statistics and death records) to the total number of dementia diagnoses identified.** This highlights the large contribution made by GP records to total numbers of diagnoses. All the estimates were based on data acquired by UK Biobank at the 2021 when the resource had access to all GP data under the UK control of patient information (COPI) notice to NHS digital for full sharing of confidential patient information during the COVID-19 pandemic (17 March 2020- 30 June 2022). Numbers of cases after that are projected based on estimated prevalence rates by age group.

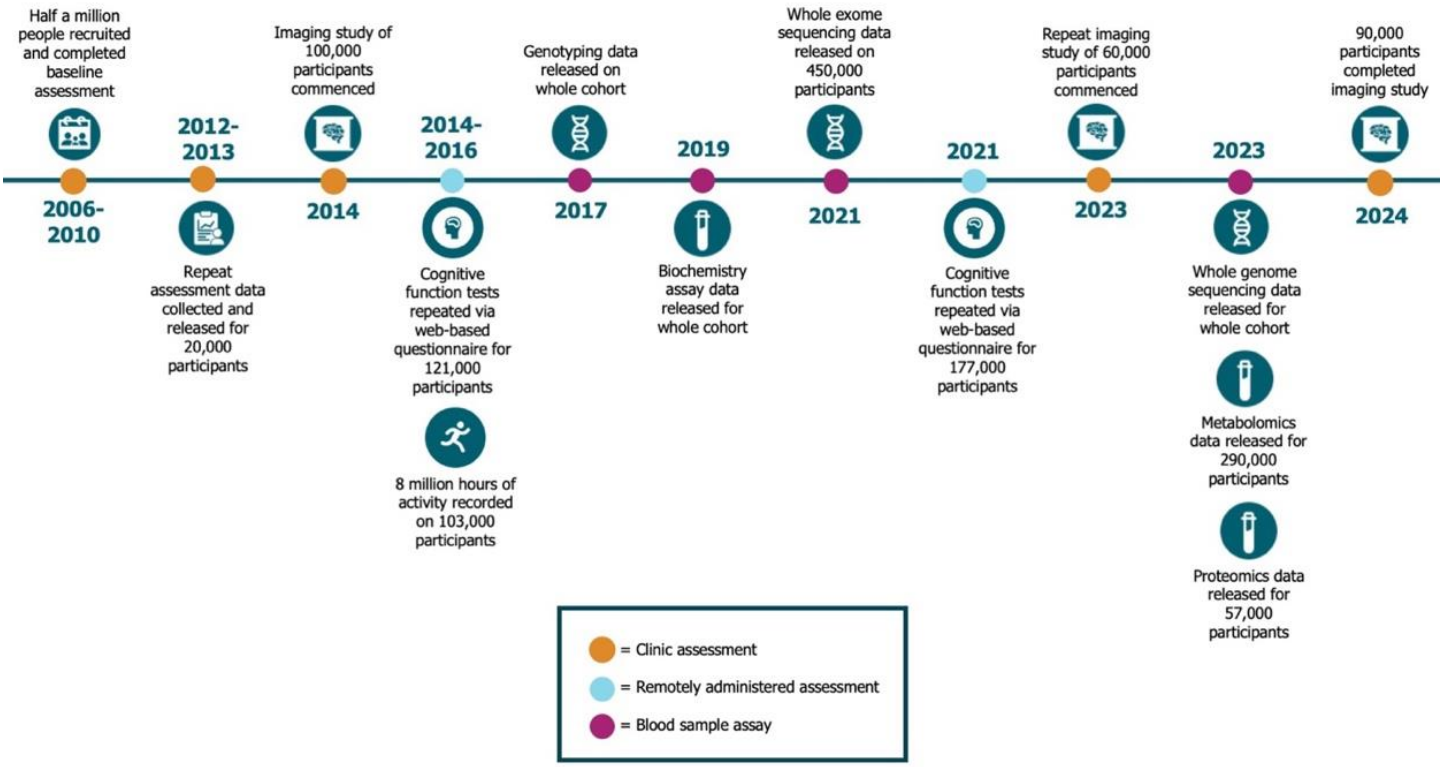

Fig. 2. Enhancements to UK Biobank. A timeline is shown for the major types of additional data for UK Biobank that were acquired by recontacting participants for additional assessments or from new assays of previously collected biosamples. Most of these enhancements have been initiated, led and funded from new resources raised from public and private sources by consortia of users of the data.

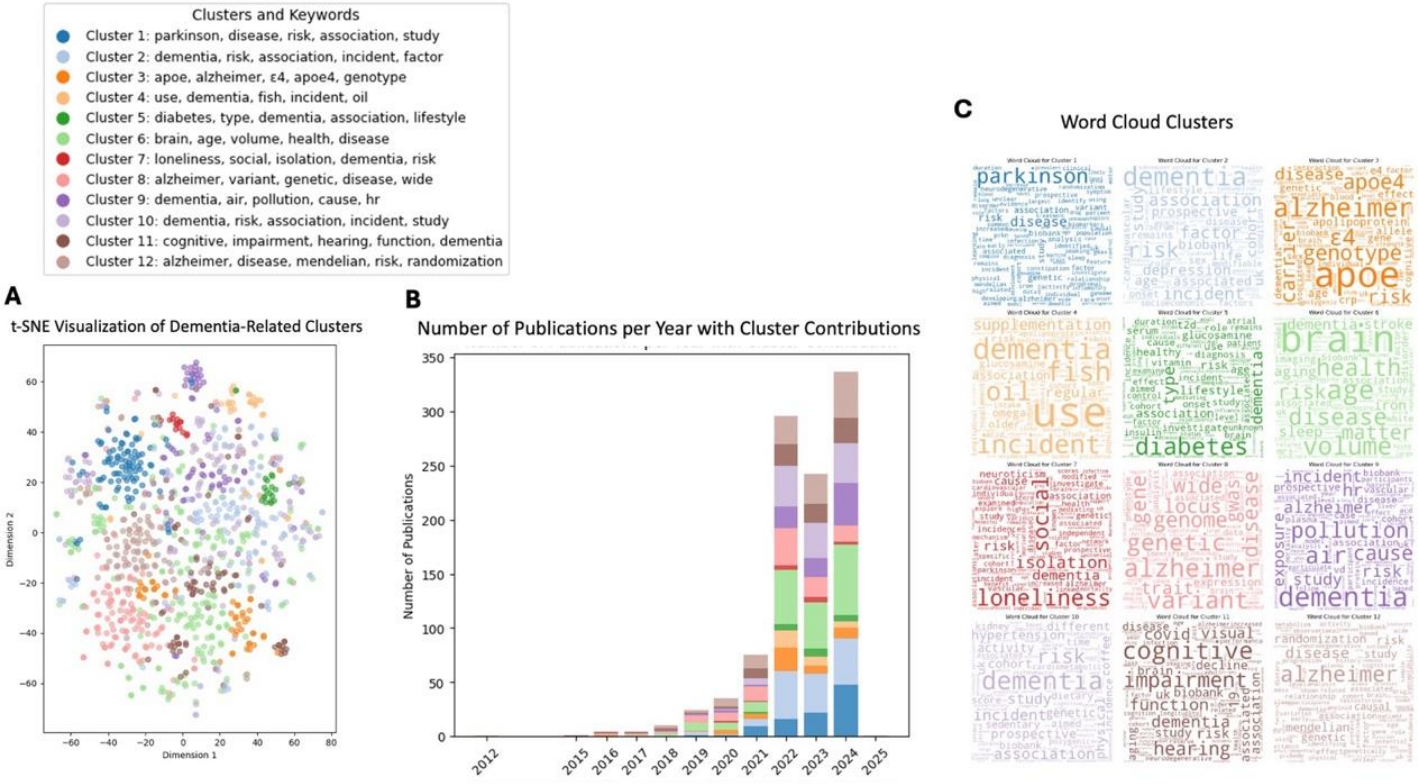

Fig. 3. Trends for dementia-related research in the UK Biobank. To illustrate the range of dementia studies conducted using UK Biobank data (2015–2024), 1,030 abstracts were retrieved from PubMed (Jan 2025) using keywords[10] related to dementia and UK Biobank. The text was processed through tokenization, stopword removal, and lemmatization. Abstracts were clustered (using latent Dirochlet allocation[154] or K-means[2]) into different topic areas which were subsequently labelled based on the most frequent and representative keywords used in each cluster. (**A) Publication key word theme clusters.** The scatter plot illustrates a t-SNE (t-distributed stochastic neighbour embedding) visualization of publication data grouped into seven thematic clusters based on keyword analysis. Each dot represents a publication and colours correspond to different sets of associated keywords. **(B) Numbers of publications per year from each of the key word clusters**. The annual growth in dementia-related publications from 2016 to 2024 is shown, grouped by cluster contributions. Each bar is color-coded to match clusters in the t-SNE plot. Major contributors to changes over time come from clusters related to Parkinson's disease, dementia risk, and Alzheimer's genetics. **(C) Key word clouds for each cluster.** Each word cloud describes a thematic cluster topic by the associated keywords. The size of each word reflects its frequency and relevance within the respective cluster.

[10] Keywords used: 'dementia', 'Alzheimer', 'Parkinson', 'Huntington', 'MCI', 'mild cognitive impairment', 'frontotemporal dementia','TBI', 'traumatic brain injury', 'Lewy body dementia', 'vascular dementia', 'neurodegenerative'

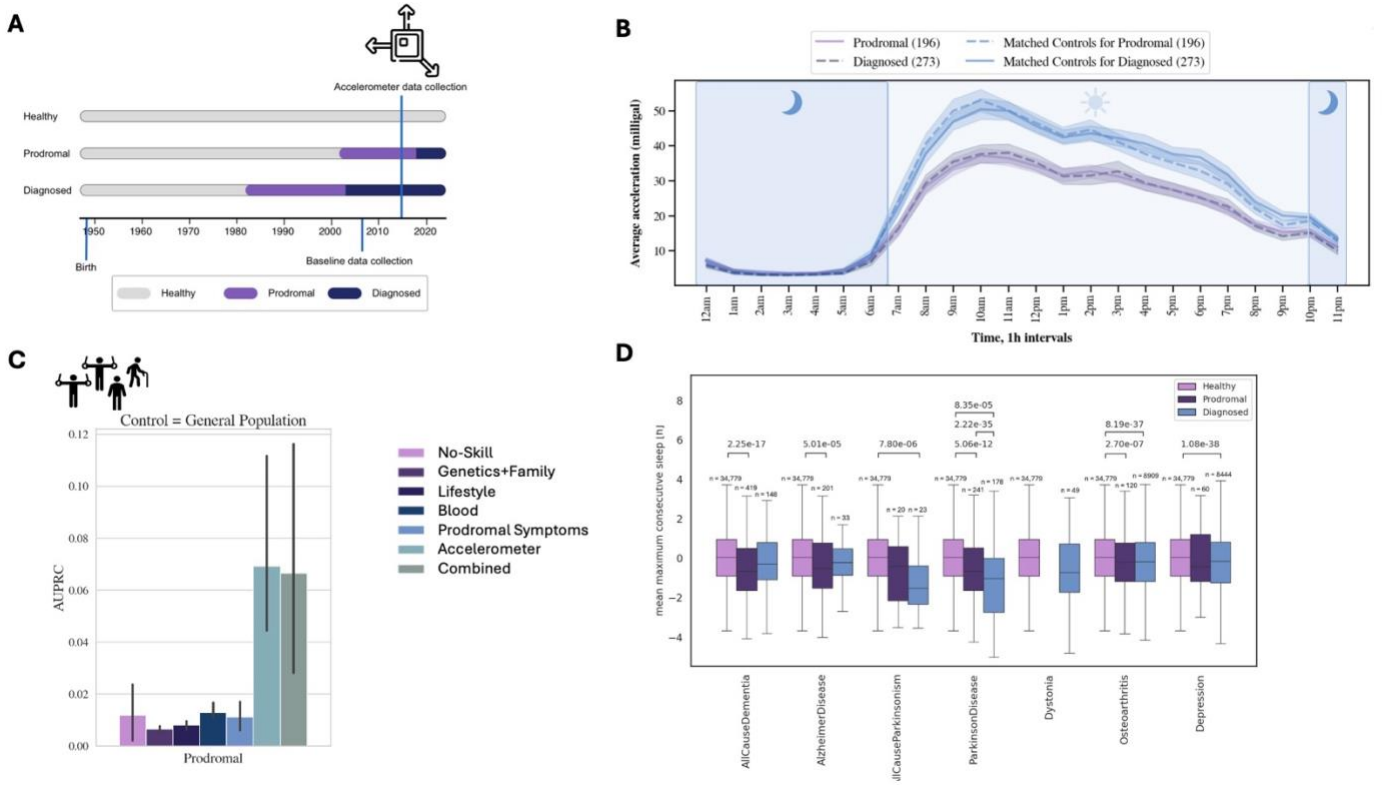

Fig. 4. Early digital detection of Parkinson's disease (PD) risk using accelerometry. **(A)** The UK Biobank collected one week of continuous wrist-worn accelerometry data from ~100,000 participants in 2015 using Axivity Ax3 devices. This was used to predict future cases of PD. Prodromal PD cases were defined as individuals without a diagnosis at the time of accelerometry assessment, but who received a PD diagnosis more than two years later. Recently diagnosed cases were those diagnosed at or shortly after accelerometry collection. Demographically matched controls who have never received a diagnosis of PD are included for comparison. **(B)** Average hourly acceleration profiles show progressively lower levels in prodromal and diagnosed individuals compared with controls, consistent with early and ongoing disease progression. **(C)** Predictive modelling using accelerometry-derived features demonstrated the ability to detect prodromal PD in the general population. Performance was evaluated using the Area Under the Precision–Recall Curve (AUPRC). Modality-specific predictors are not provided for the no-skill model, so the AUPRC reflects the level expected from random guessing. In developing the other models, modality-specific feature sets have been to the same basic model. The increment with addition of accelerometer data (green bar) highlights their predictive contribution. **(D)** Sleep-related features derived from accelerometry, adjusted for age, sex, and body mass index (BMI), revealed shorter consecutive sleep durations in prodromal PD compared to controls, with similar disruptions observed in Alzheimer's disease and broader dementia diagnoses. Modified from[35].

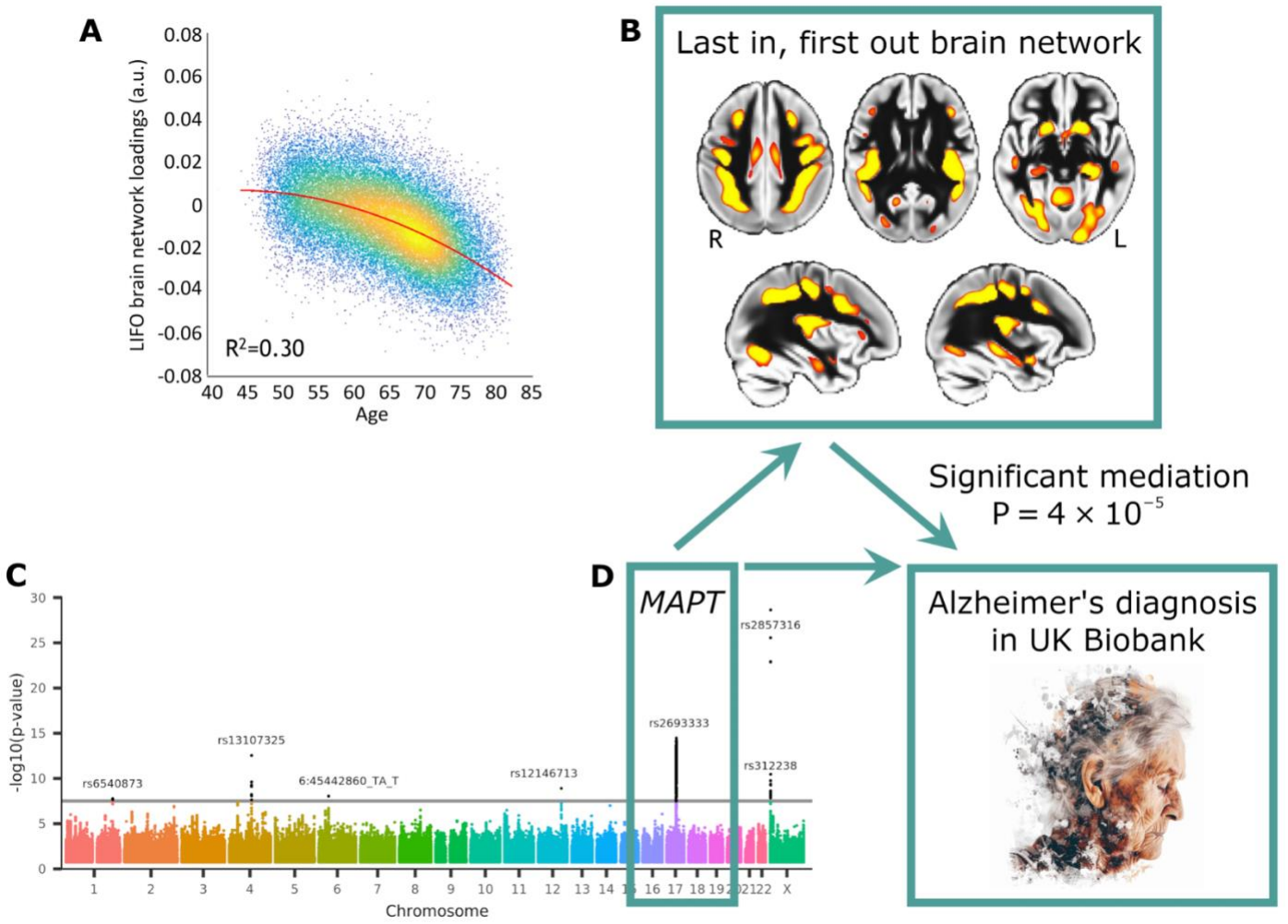

Fig. 5. Higher-order, 'transmodal' network of brain regions that degenerate earlier and faster ("last in, first out" or LIFO) than the rest of the brain spatially overlaps with the pattern of brain volume loss typical of AD as defined in[51]. **(A) LIFO loadings** (in arbitrary units), which represent the normalised grey matter volumes extracted from the LIFO brain network, decrease quadratically with age in ~40,000 UK Biobank participants ($R^2 = 0.30$, $P < 2.23 \times 10^{-308}$). **(B) The spatial map of the LIFO network** for structural images from these UK Biobank participants (in red-yellow, thresholded at $Z > 4$ for visualisation). **(C) Manhattan plot of the seven clusters in the LIFO brain network** highlights a significant association with the MAPT/KANSL1 gene cluster. **(D) LIFO brain regions mediate the effect of the *MAPT*/*KANSL1* genetic cluster on AD diagnoses** amongst participants in UK Biobank (using the lead bi-allelic variant rs2532395, the highest variant after tri-allelic rs2693333; both for the dominant and the recessive models: $P = 4 \times 10^{-5}$). The illustration of dementia (from Tim Reckmann, CC-BY 2.0) is intended to highlight a genetic link to AD through the *MAPT* locus. Modified from[51].

References

1 De Strooper, B. & Karran, E. The Cellular Phase of Alzheimer's Disease. *Cell* **164**, 603-615 (2016). https://doi.org/10.1016/j.cell.2015.12.056

2 Ikotan, A. M., Ezugwu, A.E., Abualigah,L., Abuhaija,B., Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences* **622**, 178-210 (2023).

3 Gabersek, V., Labar, P., Duche, D. J., Scherrer, J. & Michaux, L. [Sleep of enuretic children]. *Rev Neurol (Paris)* **115**, 493-497 (1966).

4 Sudlow, C. *et al.* 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**, e1001779 (2015). https://doi.org/10.1371/journal.pmed.1001779

5 Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026-1034 (2017). https://doi.org/10.1093/aje/kwx246

6 Schoeler, T. *et al.* Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav* **7**, 1216-1227 (2023). https://doi.org/10.1038/s41562-023-01579-9

7 Batty, G. D., Gale, C. R., Kivimaki, M., Deary, I. J. & Bell, S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ* **368**, m131 (2020). https://doi.org/10.1136/bmj.m131

8 van Alten, S., Domingue, B. W., Faul, J., Galama, T. & Marees, A. T. Reweighting UK Biobank corrects for pervasive selection bias due to volunteering. *Int J Epidemiol* **53** (2024). https://doi.org/10.1093/ije/dyae054

9 Mekli, K. *et al.* Investigation of the association between the antibody responses to neurotropic viruses and dementia outcomes in the UK Biobank. *PloS one* **17**, e0274872 (2022). https://doi.org/10.1371/journal.pone.0274872

10 Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PloS one* **15**, e0231627 (2020). https://doi.org/10.1371/journal.pone.0231627

11 Calvin, C. M. *et al.* Predicting incident dementia 3-8 years after brief cognitive tests in the UK Biobank prospective study of 500,000 people. *Alzheimers Dement* **15**, 1546-1557 (2019). https://doi.org/10.1016/j.jalz.2019.07.014

12 Wilkinson, T. *et al.* Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol* **34**, 557-565 (2019). https://doi.org/10.1007/s10654-019-00499-1

13 Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using family history of disease. *Nat Genet* **49**, 325-331 (2017). https://doi.org/10.1038/ng.3766

14 Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Transl Psychiatry* **8**, 99 (2018). https://doi.org/10.1038/s41398-018-0150-6

15 Wightman, D. P. *et al.* Author Correction: A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1722 (2021). https://doi.org/10.1038/s41588-021-00977-x

16 Escott-Price, V. & Hardy, J. Genome-wide association studies for Alzheimer's disease: bigger is not always better. *Brain Commun* **4**, fcac125 (2022). https://doi.org/10.1093/braincomms/fcac125

17 Wu, Y. *et al.* Pervasive biases in proxy genome-wide association studies based on parental history of Alzheimer's disease. *Nat Genet* **56**, 2696-2703 (2024). https://doi.org/10.1038/s41588-024-01963-9
This report uses the UK Biobank data in conjunction with that from other studies to define specific limitations of proxy diagnoses of AD for GWAS studies, an approach based on leveraging the comprehensive family history data that has been applied to mitigate limitations in numbers of clinically expressed AD cases in the relatively young participant group.

18 Grotzinger, A. D., Fuente, J., Prive, F., Nivard, M. G. & Tucker-Drob, E. M. Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution. *Biol Psychiatry* **93**, 29-36 (2023). https://doi.org/10.1016/j.biopsych.2022.05.029

19 Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* **51**, 404-413 (2019). https://doi.org/10.1038/s41588-018-0311-9

20 Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412-436 (2022). https://doi.org/10.1038/s41588-022-01024-z
This study, the latest GWAS of AD, involves a large-scale meta-analysis of UK Biobank diagnoses by proxy and with case-control data from other cohorts to discover the most comprehensive set

21 Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* **53**, 392-402 (2021). https://doi.org/10.1038/s41588-020-00776-w

22 Wightman, D. P., Savage, J. E., de Leeuw, C. A., Jansen, I. E. & Posthuma, D. Rare variant aggregation in 148,508 exomes identifies genes associated with proxy dementia. *Scientific reports* **13**, 2179 (2023). https://doi.org/10.1038/s41598-023-29108-8

23 Zhang, Y. R. *et al.* Whole exome sequencing analyses identified novel genes for Alzheimer's disease and related dementia. *Alzheimers Dement* **20**, 7062-7078 (2024). https://doi.org/10.1002/alz.14181

24 Neuner, S. M., Tcw, J. & Goate, A. M. Genetic architecture of Alzheimer's disease. *Neurobiology of disease* **143**, 104976 (2020). https://doi.org/10.1016/j.nbd.2020.104976

25 Raichlen, D. A. *et al.* Sedentary Behavior and Incident Dementia Among Older Adults. *JAMA* **330**, 934-940 (2023). https://doi.org/10.1001/jama.2023.15231

26 Zhong, Q. *et al.* The independent and joint association of accelerometer-measured physical activity and sedentary time with dementia: a cohort study in the UK Biobank. *Int J Behav Nutr Phys Act* **20**, 59 (2023). https://doi.org/10.1186/s12966-023-01464-8

27 Folley, S., Zhou, A., Llewellyn, D. J. & Hypponen, E. Physical Activity, APOE Genotype, and Cognitive Decline: Exploring Gene-Environment Interactions in the UK Biobank. *J Alzheimers Dis* **71**, 741-750 (2019). https://doi.org/10.3233/JAD-181132

28 Spencer, F. S. E. *et al.* The Relationship Between Physical Activity and Non-Modifiable Risk Factors on Alzheimer's Disease and Brain Health Markers: A UK Biobank Study. *J Alzheimers Dis* **101**, 1029-1042 (2024). https://doi.org/10.3233/JAD-240269

29 Petermann-Rocha, F. *et al.* Dose-response association between device-measured physical activity and incident dementia: a prospective study from UK Biobank. *BMC Med* **19**, 305 (2021). https://doi.org/10.1186/s12916-021-02172-5

30 Del Pozo Cruz, B., Ahmadi, M., Naismith, S. L. & Stamatakis, E. Association of Daily Step Count and Intensity With Incident Dementia in 78 430 Adults Living in the UK. *JAMA neurology* **79**, 1059-1063 (2022). https://doi.org/10.1001/jamaneurol.2022.2672

31 Liu, Y. *et al.* Associations between accelerometer-measured circadian rest-activity rhythm, brain structural and genetic mechanisms, and dementia. *Psychiatry Clin Neurosci* **78**, 393-404 (2024). https://doi.org/10.1111/pcn.13671

32 Winer, J. R. *et al.* Impaired 24-h activity patterns are associated with an increased risk of Alzheimer's disease, Parkinson's disease, and cognitive decline. *Alzheimers Res Ther* **16**, 35 (2024). https://doi.org/10.1186/s13195-024-01411-0

33 Yiallourou, S. R. *et al.* Association of the Sleep Regularity Index With Incident Dementia and Brain Volume. *Neurology* **102**, e208029 (2024). https://doi.org/10.1212/WNL.0000000000208029

34 Haghayegh, S. *et al.* Association of Rest-Activity Rhythm and Risk of Developing Dementia or Mild Cognitive Impairment in the Middle-Aged and Older Population: Prospective Cohort Study. *JMIR Public Health Surveill* **10**, e55211 (2024). https://doi.org/10.2196/55211

35 Schalkamp, A. K., Peall, K. J., Harrison, N. A. & Sandor, C. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. *Nat Med* **29**, 2048-2056 (2023). https://doi.org/10.1038/s41591-023-02440-2
This study demonstrates that reduced movement measured via wrist-worn accelerometers can predict Parkinson's disease (PD) up to seven years before clinical diagnosis in the general population and outperform genetic, biochemical, and symptom-based models, highlighting its potential as a low-cost, scalable digital biomarker for early PD detection.

36 Borghammer, P., Okkels, N. & Weintraub, D. Parkinson's Disease and Dementia with Lewy Bodies: One and the Same. *J Parkinsons Dis* **14**, 383-397 (2024). https://doi.org/10.3233/JPD-240002

37 Littlejohns, T. J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun* **11**, 2624 (2020). https://doi.org/10.1038/s41467-020-15948-9

38 Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience* **19**, 1523-1536 (2016). https://doi.org/10.1038/nn.4393

39 Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400-424 (2018). https://doi.org/10.1016/j.neuroimage.2017.10.034

40 Schuff, N. *et al.* MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain : a journal of neurology* **132**, 1067-1077 (2009). https://doi.org/10.1093/brain/awp007

41 Huang, S. Y. *et al.* Glymphatic system dysfunction predicts amyloid deposition, neurodegeneration, and clinical progression in Alzheimer's disease. *Alzheimers Dement* **20**, 3251-3269 (2024). https://doi.org/10.1002/alz.13789

42 Libedinsky, I. *et al.* Polyconnectomic Scoring of Functional Connectivity Patterns Across Eight Neuropsychiatric and Three Neurodegenerative Disorders. *Biol Psychiatry* **97**, 1045-1058 (2025). https://doi.org/10.1016/j.biopsych.2024.10.007

43 Wang, C. *et al.* Phenotypic and genetic associations of quantitative magnetic susceptibility in UK Biobank brain imaging. *Nature neuroscience* **25**, 818-831 (2022). https://doi.org/10.1038/s41593-022-01074-w
This report describes the creation of brain magnetic susceptibility maps for UK Biobank participants in the imaging sub-study and how these quantitative measures of iron deposition then could be used for a GWAS to discover genes responsible for population variation in this trait related to dementia pathology.

44 Ravanfar, P. *et al.* Systematic Review: Quantitative Susceptibility Mapping (QSM) of Brain Iron Profile in Neurodegenerative Diseases. *Front Neurosci* **15**, 618435 (2021). https://doi.org/10.3389/fnins.2021.618435

45 Elliott, C. *et al.* Slowly expanding/evolving lesions as a magnetic resonance imaging marker of chronic active multiple sclerosis lesions. *Multiple sclerosis (Houndmills, Basingstoke, England)*, 1352458518814117 (2018). https://doi.org/10.1177/1352458518814117

46 Smith, S. M. *et al.* An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature neuroscience* **24**, 737-745 (2021). https://doi.org/10.1038/s41593-021-00826-4

47 Buto, P. T. *et al.* Genetic risk score for Alzheimer's disease predicts brain volume differences in mid and late life in UK biobank participants. *Alzheimers Dement* **20**, 1978-1987 (2024). https://doi.org/10.1002/alz.13610

48 Wang, Z. *et al.* Bidirectional two-sample Mendelian randomization analyses support causal relationships between structural and diffusion imaging-derived phenotypes and the risk of major neurodegenerative diseases. *Transl Psychiatry* **14**, 215 (2024). https://doi.org/10.1038/s41398-024-02939-3

49 Wen, J. *et al.* Genetic and clinical correlates of two neuroanatomical AI dimensions in the Alzheimer's disease continuum. *Transl Psychiatry* **14**, 420 (2024). https://doi.org/10.1038/s41398-024-03121-5

50 Casanova, F. *et al.* Iron and risk of dementia: Mendelian randomisation analysis in UK Biobank. *J Med Genet* **61**, 435-442 (2024). https://doi.org/10.1136/jmg-2023-109295

51 Manuello, J. *et al.* The effects of genetic and modifiable risk factors on brain regions vulnerable to ageing and disease. *Nat Commun* **15**, 2576 (2024). https://doi.org/10.1038/s41467-024-46344-2
This study integrated data from UK Biobank brain imaging with genotypes and 161 of the associated data fields describing modifiable risk factors for dementia to explore their relative contributions to differences in the structure and volume of specific regions of the brain across the population, discovering evidence that those regions last to develop appeared most vulnerable to degeneration with greater risk factor loads.

52 Dove, A. *et al.* High cognitive reserve attenuates the risk of dementia associated with cardiometabolic diseases. *Alzheimers Res Ther* **16**, 161 (2024). https://doi.org/10.1186/s13195-024-01528-2

53 Jin, J., Sommerlad, A. & Mukadam, N. Association between adult education, brain volume and dementia risk: longitudinal cohort study of UK Biobank participants. *Geroscience* **47**, 903-913 (2025). https://doi.org/10.1007/s11357-024-01285-y

54 Xiong, L. Y. *et al.* Latent profiles of modifiable dementia risk factors in later midlife: relationships with incident dementia, cognition, and neuroimaging outcomes. *Mol Psychiatry* **30**, 450-460 (2025). https://doi.org/10.1038/s41380-024-02685-4

55 Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329-338 (2023). https://doi.org/10.1038/s41586-023-06592-6

56 Gadd, D. A. *et al.* Blood protein assessment of leading incident diseases and mortality in the UK Biobank. *Nat Aging* **4**, 939-948 (2024). https://doi.org/10.1038/s43587-024-00655-7

57 You, J. *et al.* Plasma proteomic profiles predict individual future health risk. *Nat Commun* **14**, 7817 (2023). https://doi.org/10.1038/s41467-023-43575-7

58 Guo, Y. *et al.* Plasma proteomic profiles predict future dementia in healthy adults. *Nat Aging* **4**, 247-260 (2024). https://doi.org/10.1038/s43587-023-00565-0
This study, along with several others, leverage the unprecedented scale of the UK Biobank's Pharma Proteomics Project data to train high-performing predictive models for dementia and other disease risks.

59 Zhang, Y. *et al.* Large-scale proteomic analyses of incident Alzheimer's disease reveal new pathophysiological insights and potential therapeutic targets. *Mol Psychiatry* **30**, 2347-2361 (2025). https://doi.org/10.1038/s41380-024-02840-x

60 Argentieri, M. A. *et al.* Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nat Med* **30**, 2450-2460 (2024). https://doi.org/10.1038/s41591-024-03164-7

61 Argentieri, M. A. *et al.* Integrating the environmental and genetic architectures of aging and mortality. *Nat Med* **31**, 1016-1025 (2025). https://doi.org/10.1038/s41591-024-03483-9

62 Beydoun, M. A. *et al.* Plasma proteomic biomarkers and the association between poor cardiovascular health and incident dementia: The UK Biobank study. *Brain, behavior, and immunity* **119**, 995-1007 (2024). https://doi.org/10.1016/j.bbi.2024.05.005

63 Julkunen, H. *et al.* Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun* **14**, 604 (2023). https://doi.org/10.1038/s41467-023-36231-7

64 Qiang, Y. X. *et al.* Plasma metabolic profiles predict future dementia and dementia subtypes: a prospective analysis of 274,160 participants. *Alzheimers Res Ther* **16**, 16 (2024). https://doi.org/10.1186/s13195-023-01379-3

65 Huang, S. Y. *et al.* Investigating Causal Relations Between Circulating Metabolites and Alzheimer's Disease: A Mendelian Randomization Study. *J Alzheimers Dis* **87**, 463-477 (2022). https://doi.org/10.3233/JAD-220050

66 Dehghan, A. *et al.* Metabolome-wide association study on ABCA7 indicates a role of ceramide metabolism in Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **119**, e2206083119 (2022). https://doi.org/10.1073/pnas.2206083119

67 Gong, J., Harris, K., Peters, S. A. E. & Woodward, M. Serum lipid traits and the risk of dementia: A cohort study of 254,575 women and 214,891 men in the UK Biobank. *EClinicalMedicine* **54**, 101695 (2022). https://doi.org/10.1016/j.eclinm.2022.101695

68 Livingston, G. *et al.* Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet (London, England)* **396**, 413-446 (2020). https://doi.org/10.1016/S0140-6736(20)30367-6

69 Livingston, G. *et al.* Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet (London, England)* **404**, 572-628 (2024). https://doi.org/10.1016/S0140-6736(24)01296-0

70 Lourida, I. *et al.* Association of Lifestyle and Genetic Risk With Incidence of Dementia. *JAMA* **322**, 430-437 (2019). https://doi.org/10.1001/jama.2019.9879

71 Floud, S. *et al.* Body mass index, diet, physical inactivity, and the incidence of dementia in 1 million UK women. *Neurology* **94**, e123-e132 (2020). https://doi.org/10.1212/WNL.0000000000008779

72 Sabia, S. *et al.* Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *BMJ* **357**, j2709 (2017). https://doi.org/10.1136/bmj.j2709

73 Richmond, R. C. & Davey Smith, G. Mendelian Randomization: Concepts and Scope. *Cold Spring Harb Perspect Med* **12** (2022). https://doi.org/10.1101/cshperspect.a040501

74 Henry, A. *et al.* The relationship between sleep duration, cognition and dementia: a Mendelian randomization study. *Int J Epidemiol* **48**, 849-860 (2019). https://doi.org/10.1093/ije/dyz071
75 Yuan, S. *et al.* Sleep duration, genetic susceptibility, and Alzheimer's disease: a longitudinal UK Biobank-based study. *BMC Geriatr* **22**, 638 (2022). https://doi.org/10.1186/s12877-022-03298-8
76 Barberger-Gateau, P. *et al.* Dietary patterns and risk of dementia: the Three-City cohort study. *Neurology* **69**, 1921-1930 (2007). https://doi.org/10.1212/01.wnl.0000278116.37320.52
77 Cornelis, M. C., Agarwal, P., Holland, T. M. & van Dam, R. M. MIND Dietary Pattern and Its Association with Cognition and Incident Dementia in the UK Biobank. *Nutrients* **15** (2022). https://doi.org/10.3390/nu15010032
78 Liu, X. *et al.* Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) study: Rationale, design and baseline characteristics of a randomized control trial of the MIND diet on cognitive decline. *Contemp Clin Trials* **102**, 106270 (2021). https://doi.org/10.1016/j.cct.2021.106270
79 Shannon, O. M. *et al.* Mediterranean diet adherence is associated with lower dementia risk, independent of genetic predisposition: findings from the UK Biobank prospective cohort study. *BMC Med* **21**, 81 (2023). https://doi.org/10.1186/s12916-023-02772-3
80 Ma, H., Zhou, T., Li, X., Heianza, Y. & Qi, L. Use of fish oil supplements is differently related to incidence of all-cause and vascular dementia among people with the distinct APOE epsilon4 dosage. *Clin Nutr* **41**, 731-736 (2022). https://doi.org/10.1016/j.clnu.2022.01.019
81 Wang, B. *et al.* Association between coffee and tea consumption and the risk of dementia in individuals with hypertension: a prospective cohort study. *Scientific reports* **14**, 21063 (2024). https://doi.org/10.1038/s41598-024-71426-y
82 Zhou, A. *et al.* Habitual coffee consumption and cognitive function: a Mendelian randomization meta-analysis in up to 415,530 participants. *Scientific reports* **8**, 7526 (2018). https://doi.org/10.1038/s41598-018-25919-2

83 Duperron, M. G. *et al.* Genomics of perivascular space burden unravels early mechanisms of cerebral small vessel disease. *Nat Med* **29**, 950-962 (2023). https://doi.org/10.1038/s41591-023-02268-w
UK Biobank imaging and genetic (genotyping, whole exome and whole genome sequencing) data made the largest contribution to the aggregated data used for genetic association studies of determinants of perivascular space structure, an emerging and previously unexplored imaging correlate of cSVD.

84 Sipila, P. N. *et al.* Hospital-treated infectious diseases and the risk of dementia: a large, multicohort, observational study with a replication cohort. *Lancet Infect Dis* **21**, 1557-1567 (2021). https://doi.org/10.1016/S1473-3099(21)00144-4
85 Beydoun, M. A. *et al.* Hospital-treated prevalent infections, the plasma proteome and incident dementia among UK older adults. *iScience* **26**, 108526 (2023). https://doi.org/10.1016/j.isci.2023.108526

86 Levine, K. S. *et al.* Virus exposure and neurodegenerative disease risk across national biobanks. *Neuron* **111**, 1086-1093 e1082 (2023). https://doi.org/10.1016/j.neuron.2022.12.029
After combining data from UK and FinnGen biobank datasets, this study identified 45 associations of viral exposures with neurodegenerative disease, highlighting a potential mechanistic relationship.

87 Zhao, H. *et al.* Prospective cohort study evaluating the association between influenza vaccination and neurodegenerative diseases. *NPJ Vaccines* **9**, 51 (2024). https://doi.org/10.1038/s41541-024-00841-z
88 Oh, H. S. *et al.* Plasma proteomics in the UK Biobank reveals youthful brains and immune systems promote healthspan and longevity. *bioRxiv* (2024). https://doi.org/10.1101/2024.06.07.597771
89 Allen, N. E. *et al.* Prospective study design and data analysis in UK Biobank. *Science translational medicine* **16**, eadf4428 (2024). https://doi.org/10.1126/scitranslmed.adf4428
90 Giovane, M. D. *et al.* Remote cognitive tests predict neurodegenerative biomarkers in the Insight 46 cohort. *Alzheimers Dement* **21**, e14572 (2025). https://doi.org/10.1002/alz.14572
91 Wu, X., Peng, C., Cheng, Q. Sex-adjusted blood biomarkers differentiate AD from LATE. *Alzheimer's Dement.* **19**, 060446 (2023).
92 Nelson, P. T., Schneider, J. A., Jicha, G. A., Duong, M. T. & Wolk, D. A. When Alzheimer's is LATE: Why Does it Matter? *Annals of neurology* **94**, 211-222 (2023). https://doi.org/10.1002/ana.26711

93 Mc Ardle, R. *et al.* Factors That Influence Habitual Activity in Mild Cognitive Impairment and Dementia. *Gerontology* **66**, 197-208 (2020). https://doi.org/10.1159/000502288

94 Heremans, E. R. M. *et al.* Wearable sleep recording augmented by artificial intelligence for Alzheimer's disease screening. *NPJ Aging* **11**, 34 (2025). https://doi.org/10.1038/s41514-025-00219-y

95 Parra, K. L., Alexander, G. E., Raichlen, D. A., Klimentidis, Y. C. & Furlong, M. A. Exposure to air pollution and risk of incident dementia in the UK Biobank. *Environ Res* **209**, 112895 (2022). https://doi.org/10.1016/j.envres.2022.112895

96 Wallace, H. M. The development of UK Biobank: Excluding scientific controversy from ethical debate. *Critical Public Health* **15**, 323–333 (2006).

97 Yang, L. *et al.* Depression, Depression Treatments, and Risk of Incident Dementia: A Prospective Cohort Study of 354,313 Participants. *Biol Psychiatry* **93**, 802-809 (2023). https://doi.org/10.1016/j.biopsych.2022.08.026

98 Stevenson, J. S., Clifton, L., Kuzma, E. & Littlejohns, T. J. Speech-in-noise hearing impairment is associated with an increased risk of incident dementia in 82,039 UK Biobank participants. *Alzheimers Dement* **18**, 445-456 (2022). https://doi.org/10.1002/alz.12416

99 Zhang, L. *et al.* The Impact of Testosterone on Alzheimer's Disease Are Mediated by Lipid Metabolism and Obesity: A Mendelian Randomization Study. *J Prev Alzheimers Dis* **11**, 507-513 (2024). https://doi.org/10.14283/jpad.2023.116

100 Shang, X. *et al.* The Association of Age at Diagnosis of Hypertension With Brain Structure and Incident Dementia in the UK Biobank. *Hypertension* **78**, 1463-1474 (2021). https://doi.org/10.1161/HYPERTENSIONAHA.121.17608

101 Sproviero, W. *et al.* High Blood Pressure and Risk of Dementia: A Two-Sample Mendelian Randomization Study in the UK Biobank. *Biol Psychiatry* **89**, 817-824 (2021). https://doi.org/10.1016/j.biopsych.2020.12.015

102 Desai, R., Heller, J., Wada, R., Anis, M., Afnan, M., Cheong, C., Zhang, S., Deo, F., Tam, M., Anderson, E., Stott, J.,Charlesworth, G., Yarmolinsky, J.W., Elliott, P., Chadeau-Hyam, M., Zuber, V. Hypertension and Dementia: A Mendelian Randomisation Study Assessing Potential Causal Relationships with All-Cause Dementia, Alzheimer's Disease and Vascular Dementia in the UK Biobank. *medRxiv*, 2025.2003.2012.25323830 (2025). https://doi.org/doi.org/10.1101/2025.03.12.25323830

103 Deng, Y. T. *et al.* Association of life course adiposity with risk of incident dementia: a prospective cohort study of 322,336 participants. *Mol Psychiatry* **27**, 3385-3395 (2022). https://doi.org/10.1038/s41380-022-01604-9

104 Chen, H. *et al.* Age- and sex-specific modifiable risk factor profiles of dementia: evidence from the UK Biobank. *Eur J Epidemiol* **38**, 83-93 (2023). https://doi.org/10.1007/s10654-022-00952-8

105 Liu, J. *et al.* The impact of maternal smoking during pregnancy and the age of smoking initiation on incident dementia: A prospective cohort study. *Alzheimers Dement* **20**, 4066-4079 (2024). https://doi.org/10.1002/alz.13854

106 Zhang, N. *et al.* Interaction between genetic predisposition, smoking, and dementia risk: a population-based cohort study. *Scientific reports* **11**, 12953 (2021). https://doi.org/10.1038/s41598-021-92304-x

107 Shen, C. & Feng, J. Author Response: Associations of Social Isolation and Loneliness With Later Dementia. *Neurology* **99**, 1014 (2022). https://doi.org/10.1212/WNL.0000000000201564

108 Zhou, C. *et al.* Sex-specific associations between diabetes and dementia: the role of age at onset of disease, insulin use and complications. *Biol Sex Differ* **14**, 9 (2023). https://doi.org/10.1186/s13293-023-00491-1

109 Hendriks, S. *et al.* Risk Factors for Young-Onset Dementia in the UK Biobank. *JAMA neurology* **81**, 134-142 (2024). https://doi.org/10.1001/jamaneurol.2023.4929

110 Jeong, S., Lin, L., Leone, A. P. & Hsu, Y. H. Type 2 diabetes and late-onset Alzheimer's disease and related dementia: A longitudinal cohort study integrating polygenic risk score. *J Alzheimers Dis* **105**, 107-119 (2025). https://doi.org/10.1177/13872877251326107

111 Tari, B., Kunzi, M., Pflanz, C. P., Raymont, V. & Bauermeister, S. Education is power: preserving cognition in the UK biobank. *Front Public Health* **11**, 1244306 (2023). https://doi.org/10.3389/fpubh.2023.1244306

112 Zhao, Y. L. *et al.* Variables associated with cognitive function: an exposome-wide and mendelian randomization analysis. *Alzheimers Res Ther* **17**, 13 (2025). https://doi.org/10.1186/s13195-025-01670-5

113 Yang, W. *et al.* Association of cognitive reserve with the risk of dementia in the UK Biobank: role of polygenic factors. *Br J Psychiatry* **224**, 213-220 (2024). https://doi.org/10.1192/bjp.2024.13
114 Zhao, A. *et al.* Objectively Measured Physical Activity Using Wrist-Worn Accelerometers as a Predictor of Incident Alzheimer's Disease in the UK Biobank. *J Gerontol A Biol Sci Med Sci* **80** (2025). https://doi.org/10.1093/gerona/glae287
115 Huang, X., Tan, C. S., Kandiah, N. & Hilal, S. Association of physical activity with dementia and cognitive decline in UK Biobank. *Alzheimers Dement (Amst)* **15**, e12476 (2023). https://doi.org/10.1002/dad2.12476
116 Daviet, R. *et al.* Associations between alcohol consumption and gray and white matter volumes in the UK Biobank. *Nat Commun* **13**, 1175 (2022). https://doi.org/10.1038/s41467-022-28735-5
117 Evangelou, E. *et al.* Alcohol consumption in the general population is associated with structural changes in multiple organ systems. *Elife* **10** (2021). https://doi.org/10.7554/eLife.65325
118 Zheng, L. *et al.* Association between alcohol consumption and incidence of dementia in current drinkers: linear and non-linear mendelian randomization analysis. *EClinicalMedicine* **76**, 102810 (2024). https://doi.org/10.1016/j.eclinm.2024.102810
119 Zhang, S. *et al.* Joint Exposure to Multiple Air Pollutants, Genetic Susceptibility, and Incident Dementia: A Prospective Analysis in the UK Biobank Cohort. *Int J Public Health* **69**, 1606868 (2024). https://doi.org/10.3389/ijph.2024.1606868
120 Wang, B. *et al.* Association between air pollution and lifestyle with the risk of developing mild cognitive impairment and dementia in individuals with cardiometabolic diseases. *Scientific reports* **15**, 2089 (2025). https://doi.org/10.1038/s41598-024-83607-w

121 Ge, R., Wang, Y., Zhang, Z., Sun, H. & Chang, J. Association of long-term exposure to various ambient air pollutants, lifestyle, and genetic predisposition with incident cognitive impairment and dementia. *BMC Public Health* **24**, 179 (2024). https://doi.org/10.1186/s12889-024-17702-y
This study applied data made available in UK Biobank from a land-use regression model that estimated exposures to $PM_{2.5}$, PM_{10}, $PM_{2.5-10}$, NO_2, and NO across the country in conjunction with residential coordinates provided for participants and a variety of health related data to provide evidence both that exposure to air pollution contributes to cognitive impairment and dementia and that a healthy lifestyle can modify this risk.

122 de Rojas, I. *et al.* Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat Commun* **12**, 3417 (2021). https://doi.org/10.1038/s41467-021-22491-8
123 Leonenko, G. *et al.* Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nat Commun* **12**, 4506 (2021). https://doi.org/10.1038/s41467-021-24082-z
124 Escott-Price, V., Myers, A. J., Huentelman, M. & Hardy, J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. *Annals of neurology* **82**, 311-314 (2017). https://doi.org/10.1002/ana.24999
125 Harrison, J. R. *et al.* Pathway-specific polygenic scores for Alzheimer's disease are associated with changes in brain structure in younger and older adults. *Brain Commun* **5**, fcad229 (2023). https://doi.org/10.1093/braincomms/fcad229
126 Yang, H. S. *et al.* Cell-type-specific Alzheimer's disease polygenic risk scores are associated with distinct disease processes in Alzheimer's disease. *Nat Commun* **14**, 7659 (2023). https://doi.org/10.1038/s41467-023-43132-2
127 Stevenson-Hoare, J. *et al.* Plasma biomarkers and genetics in the diagnosis and prediction of Alzheimer's disease. *Brain : a journal of neurology* **146**, 690-699 (2023). https://doi.org/10.1093/brain/awac128
128 Debette, S., Schilling, S., Duperron, M. G., Larsson, S. C. & Markus, H. S. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. *JAMA neurology* **76**, 81-94 (2019). https://doi.org/10.1001/jamaneurol.2018.3122
129 Psaty, B. M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* **2**, 73-80 (2009). https://doi.org/10.1161/CIRCGENETICS.108.829747
130 Raffeld, M. R., Debette, S. & Woo, D. International Stroke Genetics Consortium Update. *Stroke* **47**, 1144-1145 (2016). https://doi.org/10.1161/STROKEAHA.116.012682
131 Bordes, C., Sargurupremraj, M., Mishra, A. & Debette, S. Genetics of common cerebral small vessel disease. *Nature reviews. Neurology* **18**, 84-101 (2022). https://doi.org/10.1038/s41582-021-00592-8

132 Persyn, E. *et al.* Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun* **11**, 2175 (2020). https://doi.org/10.1038/s41467-020-15932-3
133 Sargurupremraj, M. *et al.* Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun* **11**, 6285 (2020). https://doi.org/10.1038/s41467-020-19111-2
134 Rutten-Jacobs, L. C. A. *et al.* Genetic Study of White Matter Integrity in UK Biobank (N=8448) and the Overlap With Stroke, Depression, and Dementia. *Stroke* **49**, 1340-1347 (2018). https://doi.org/10.1161/STROKEAHA.118.020811
135 Mishra, A. *et al.* Gene-mapping study of extremes of cerebral small vessel disease reveals TRIM47 as a strong candidate. *Brain : a journal of neurology* **145**, 1992-2007 (2022). https://doi.org/10.1093/brain/awab432
136 Francis, C. M. *et al.* Genome-wide associations of aortic distensibility suggest causality for aortic aneurysms and brain white matter hyperintensities. *Nat Commun* **13**, 4505 (2022). https://doi.org/10.1038/s41467-022-32219-x
137 Zhao, B. *et al.* Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol Psychiatry* **26**, 3943-3955 (2021). https://doi.org/10.1038/s41380-019-0569-z
138 Zhao, B. *et al.* Common genetic variation influencing human white matter microstructure. *Science* **372** (2021). https://doi.org/10.1126/science.abf3736
139 Le Grand, Q. *et al.* Diffusion imaging genomics provides novel insight into early mechanisms of cerebral small vessel disease. *Mol Psychiatry* **29**, 3567-3579 (2024). https://doi.org/10.1038/s41380-024-02604-7

140 Huang, S. Y. *et al.* Genome-wide association study unravels mechanisms of brain glymphatic activity. *Nat Commun* **16**, 626 (2025). https://doi.org/10.1038/s41467-024-55706-9
Leveraging UK Biobank brain imaging and genetics, this report presents new mechanistic insights into potential determinants of brain glymphatic activity, an important factor in AD pathology, through GWAS of an index based on brain diffusion-tensor imaging measures along the perivascular space (ALPS).

141 Malik, R. *et al.* Genetically proxied HTRA1 protease activity and circulating levels independently predict risk of ischemic stroke and coronary artery disease. *Nat Cardiovasc Res* **3**, 701-713 (2024). https://doi.org/10.1038/s44161-024-00475-3
142 Malik, R. *et al.* Whole-exome sequencing reveals a role of HTRA1 and EGFL8 in brain white matter hyperintensities. *Brain : a journal of neurology* **144**, 2670-2682 (2021). https://doi.org/10.1093/brain/awab253
143 Debette, S. & Ihara, M. Redefining common and rare HTRA1 variants as risk factors for polyvascular disease. *Nat Cardiovasc Res* **3**, 619-621 (2024). https://doi.org/10.1038/s44161-024-00492-2

144 Duff, E. P. *et al.* Plasma proteomic evidence for increased beta-amyloid pathology after SARS-CoV-2 infection. *Nat Med* **31**, 797-806 (2025). https://doi.org/10.1038/s41591-024-03426-4
Samples of UK Biobank participants in the COVID sub-study were accessed for assays of plasma biomarkers related to brain beta amyloid pathology and well-controlled evidence for accelerated progression of these biomarkers with SARS-CoV-2 infection even in mild to moderate cases was found.

145 Harshfield, E. L. & Markus, H. S. Association of Baseline Metabolomic Profiles With Incident Stroke and Dementia and With Imaging Markers of Cerebral Small Vessel Disease. *Neurology* **101**, e489-e501 (2023). https://doi.org/10.1212/WNL.0000000000207458
146 Debette, S. *et al.* Proteogenomics in cerebrospinal fluid and plasma reveals new biological fingerprint of cerebral small vessel disease. *Res Sq* (2024). https://doi.org/10.21203/rs.3.rs-4535534/v1
147 Fan, Z. *et al.* The landscape of plasma proteomic links to human organ imaging. *medRxiv* (2025). https://doi.org/10.1101/2025.01.14.25320532
148 Knol, M. J. *et al.* Association of common genetic variants with brain microbleeds: A genome-wide association study. *Neurology* **95**, e3331-e3343 (2020). https://doi.org/10.1212/WNL.0000000000010852

149 Atkins, J. L. *et al.* Preexisting Comorbidities Predicting COVID-19 and Mortality in the UK Biobank Community Cohort. *J Gerontol A Biol Sci Med Sci* **75**, 2224-2230 (2020). https://doi.org/10.1093/gerona/glaa183
This study used UK Biobank to show that specific preexisting comorbidities including dementia were associated with high risk for COVID-19 hospitalization, not age alone.

150 Hu, Y. *et al.* COVID-19 related outcomes among individuals with neurodegenerative diseases: a cohort analysis in the UK biobank. *BMC Neurol* **22**, 15 (2022). https://doi.org/10.1186/s12883-021-02536-7

151 Yu, Y., Travaglio, M., Popovic, R., Leal, N. S. & Martins, L. M. Alzheimer's and Parkinson's Diseases Predict Different COVID-19 Outcomes: A UK Biobank Study. *Geriatrics (Basel)* **6** (2021). https://doi.org/10.3390/geriatrics6010010

152 Kuo, C. L. *et al.* APOE e4 Genotype Predicts Severe COVID-19 in the UK Biobank Community Cohort. *J Gerontol A Biol Sci Med Sci* **75**, 2231-2232 (2020). https://doi.org/10.1093/gerona/glaa131

153 Douaud, G. *et al.* SARS-CoV-2 is associated with changes in brain structure in UK Biobank. *Nature* **604**, 697-707 (2022). https://doi.org/10.1038/s41586-022-04569-5

154 Blei, D. M., g, A.Y., Jordan, M.Il. Latent Dirichlet Allocation. *J Machine Learn* **3**, 993-1022 (2003).