Risk factors for early-onset and late-onset dementia: a prospective cohort study

Katherine Giorgio, John J Stephen, Maxwell Mansolf, Elizabeth A Peterson, Alden L Gross, Emily M Briceño, Deborah A Levine, Catherine Helmer, Stéphanie Debette, Aïcha Soumaré, M Arfan Ikram, Frank J Wolters, Sudha Seshadri, Claudia L Satizabal, Jayandra Jung Himali, Lenore J Launer, David Li, Djass Mbangdadji, Donald M Lloyd-Jones, Farzaneh A Sorond, Lihui Zhao, Oscar L Lopez, Suzanne E Judd, Timothy M Hughes, Vilmundur Guðnason, Allison E Aiello, Alison Fohner, Archana Singh-Manoux, Behnam Sabayan, James S Pankow, Paul S de Vries, Kristine Yaffe, Alexa Beiser, Abbas Dehghan, Rachel Zmora, Denise Scholtens, Pamela L Lutsey, Norrina B Allen, Richard MacLehose, Sanaz Sedaghat

Summary

Background Early-onset dementia (onset before age 65 years) is an important health concern, but much of our understanding of its risk factors is inferred from studies of late-onset dementia (onset after age 65 years). We investigated associations between several demographic, clinical, and lifestyle factors with early-onset dementia and compared those estimates against their associations with late-onset dementia.

Methods Data from five community-based longitudinal cohort studies from the UK and USA were pooled and rigorously harmonised: UK Biobank, Atherosclerosis Risk in Communities Study, Framingham Heart Study, Multi-Ethnic Study of Atherosclerosis, and Whitehall II Study. Dementia was ascertained via hospitalisation and death records with or without clinical assessments according to each cohort's protocol. Risk factors included sex, self-reported race or ethnicity (Hispanic, White, Black, Asian, and Other), low education, hypertension, diabetes, obesity, hypercholesterolaemia, depression, alcohol overconsumption, smoking, and physical inactivity. Cox regression models, with age as the timescale and time-varying coefficients, were fitted to estimate hazard ratios (HRs) for early-onset dementia and late-onset dementia and to test whether the HRs differed by age of onset.

Findings In 544 442 participants, there were 807 incident early-onset dementia cases and 14 253 incident late-onset dementia cases over a median follow-up of 13·7 years (IQR 12·9–14·4). Female participants had a lower hazard of early-onset dementia compared with males (HR 0·70 [95% CI 0·61–0·80]). Black versus White race (1·61 [1·23–2·11]), grade school education or less (1·99 [1·67–2·38]), diabetes (2·45 [1·99–3·03]), depression (2·73 [2·34–3·20]), smoking (1·86 [1·56–2·22]), obesity (1·24 [1·04–1·48]), physical inactivity (1·33 [1·11–1·59]), and alcohol overconsumption (1·22 [1·01–1·47]) were independently associated with higher hazards of early-onset dementia. Hypertension stage 1 (HR 1·19 [95% CI 0·97–1·47]), hypertension stage 2 (1·16 [0·94–1·43]), and hypercholesterolaemia (1·11 [0·92–1·34]) had positive effect estimates but were not statistically significant. All risk factors had stronger associations with early-onset dementia than with late-onset dementia except race, physical inactivity, and alcohol overconsumption.

Interpretation Our findings demonstrate the importance of modifiable risk factors in the development of early-onset dementia and guide future research for identifying high-priority targets for primary prevention.

Funding US National Institutes of Health, the National Institute for Neurologic Disorders and Stroke, and the National Institute of Aging.

Introduction

Dementia is a major, growing public health concern globally.[1] For example, in the USA, a 2025 publication indicates that the lifetime risk of developing dementia after age 55 years is 42%, which is more than double previous estimates.[2] This finding translates to approximately half a million new cases in 2025 in the USA, with projections reaching one million annually by 2060.[2] Cases that occur before the age of 65 years are considered early-onset dementia.[3] This age threshold of 65 years defining early versus late-onset dementia was chosen because of the differential social and societal effect of experiencing the disease in middle age, as defined by the historical US standard retirement age.[3] Early-onset dementia is suggested to have a stronger genetic basis and can have different clinical presentations, such as the initial affected cognitive domains.[3–5]

Much of our understanding of dementia risk factors originates from studies consisting of mostly late-onset cases. It is unclear whether the risk factors for late-onset dementia are also risk factors for early-onset dementia and whether they have the same magnitude of effect. Although some work has been done to estimate the association of modifiable risk factors with early-onset dementia, findings are inconsistent, especially for cardiovascular risk factors such as hypertension and diabetes.[6–11] Additionally, these studies excluded or did not examine the burden of

Lancet Healthy Longev 2026

Published Online https://doi.org/10.1016/j.lanhl.2026.100831

Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA (K Giorgio MS, Prof J S Pankow PhD, Prof P L Lutsey PhD, Prof R MacLehose PhD, S Sedaghat PhD); **Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA** (J J Stephen MPH, E A Peterson MPH, Prof F A Sorond MD PhD, L Zhao PhD, R Zmora PhD, Prof D Scholtens PhD, Prof N B Allen PhD); **Department of Medical Social Science, Northwestern University Feinberg School of Medicine, Chicago, IL, USA** (M Mansolf PhD); **Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA** (Prof A L Gross PhD); **Department of Physical Medicine & Rehabilitation, University of Michigan Medical School, Ann Arbor, MI, USA** (E M Briceño PhD); **Department of Internal Medicine, Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, USA** (Prof D A Levine MD MPH); **Bordeaux Population Health, University of Bordeaux, INSERM U1219, Bordeaux, France** (C Helmer MD PhD, Prof S Debette MD PhD, A Soumaré PhD); **Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, Netherlands** (Prof M A Ikram MD PhD, F J Wolters PhD); **Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases,**

University of Texas Health Science Center at San Antonio, San Antonio, TX, USA (Prof S Seshadri MD, C L Satizabal PhD, J J Himali PhD); **The Framingham Heart Study, Framingham, MA, USA** (Prof S Seshadri MD, C L Satizabal PhD, J J Himali PhD, Prof D M Lloyd-Jones MD, Prof A Beiser PhD); **Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA** (J J Himali PhD, Prof A Beiser PhD); **National Institute on Aging, National Institutes of Health, Bethesda, MD, USA** (L J Launer PhD, D Li PhD, D Mbangdadji MS); **Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA** (Prof D M Lloyd-Jones MD); **Department of Neurology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA** (Prof O L Lopez MD); **Departments of Health Behavior and Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA** (Prof S E Judd PhD); **Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA** (T M Hughes PhD); **Icelandic Heart Association, Kopavoguar, Iceland** (Prof V Guðnason MD PhD); **Faculty of Medicine, University of Reykjavik, Reykjavik, Iceland** (Prof V Guðnason MD PhD); **Robert N Butler Columbia Aging Center, Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA** (Prof A E Aiello PhD); **Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA** (A Fohner PhD); **Université Paris Cité, Inserm 1153, Paris, France** (Prof A Singh-Manoux PhD); **Faculty of Brain Sciences, University College London, UK** (Prof A Singh-Manoux PhD); **Department of Neurology, Hennepin Healthcare Research Institute, Hennepin County Medical Center, Minneapolis, MN, USA** (B Sabayan MD PhD); **Department of Neurology, University of Minnesota Medical School, Minneapolis, MN, USA** (B Sabayan MD PhD);

Research in context

Evidence before this study

A systematic review of the literature using PubMed, EMBASE, and PsycINFO databases, searching all literature from Jan 1, 1990, to 14 June, 2023, in any language, was conducted to identify evidence for associations of body mass index, systolic and diastolic blood pressure, diabetes status or glucose level, lipid profile, smoking, alcohol consumption, physical activity, depression, and educational attainment with early-onset dementia. Search terms included (non-exhaustive) "dementia", "early onset", "young onset", "presenile", "body mass index", "overweight", "obesity", "hypertension", "blood pressure", "diabetes mellitus", "blood glucose", "metabolic syndrome", "hyperlipidemia", "cholesterol", "kidney disease", "renal disease", "smoking", "tobacco", "alcohol", "drinking", "substance-related disorders", "exercise", "physical activity", "sedentary behavior", "depression", "social isolation", and "educational attainment". Additional searches were conducted between June 30, 2023, and June 30, 2025 to identify new publications. Most evidence suggested detrimental associations between these risk factors and early-onset dementia; however, heterogeneity was high, precision was low, and data were sparse for most risk factors. Publications on the comparison of associations between early-onset dementia and late-onset dementia were also searched in PubMed, resulting in three publications which made this comparison for several risk factors in a general population. Evidence so far suggests that risk factor profiles might differ between early-onset dementia and late-onset dementia, and the associations tend to be stronger in early-onset dementia.

Added value of this study

This study expands on the current evidence in several ways. First, the sample consists of five harmonised community-based cohort studies. This approach improves the diversity of the sample and also provides well defined disease risk factor exposure variables based on clinical measurements rather than self-report or administrative data. Second, we provide a statistically rigorous comparison of the magnitudes of association between these risk factors and early-onset versus late-onset dementia. Finally, we are the first group, to our knowledge, to do this comparison with a wide range of risk factor exposures in the general population within dementia subtypes, alongside all-cause dementia.

Implications of all the available evidence

Our work, which builds on other published research, suggests that modifiable risk factors are important in understanding the development of early-onset dementia, and that we might be able to identify key risk factors to target for primary prevention.

early-onset dementia and associations within non-White racial or ethnic groups or by dementia subtype.[6,7,9] Furthermore, although early-onset dementia is typically defined as dementia onset before the age of 65 years, it is also of scientific interest to examine risk factors using different age cutoffs, since the current definition is shaped primarily by societal rather than clinical implications.

We harmonised and pooled data from the UK Biobank and four prospective cohort studies from the Dementia Risk Prediction Project (DRPP) to achieve a large sample size of early-onset dementia cases, and to examine whether known risk factors for late-onset dementia have similar associations with early-onset dementia. We examined associations using multiple age cutoffs for the definition of early-onset dementia and by dementia subtype. We hypothesised that some known risk factors for late-onset dementia can pose similar or stronger associations for early-onset dementia, which might have implications for risk reduction and early prevention.

Methods

Study design and participants

For this study, we used data from five community-based longitudinal cohort studies: the UK Biobank and four studies from the DRPP (Atherosclerosis Risk in Communities [ARIC] Study, Framingham Heart Study [FHS], Multi-Ethnic Study of Atherosclerosis [MESA], and Whitehall II). These cohorts include participants aged 24–86 years at baseline residing in the USA and the UK, recruited from 1948 to 2005 for FHS, 1985 to 1988 for Whitehall II, 1987 to 1989 for ARIC, 2000 to 2002 for MESA, and 2006 to 2010 for UK Biobank. Details of participants and the data collected have been documented elsewhere.[12,13] The baseline visit for the present analysis was defined as the first visit where dementia ascertainment began. This occurred at visit 1 for UK Biobank, ARIC, MESA, Whitehall II, and the FHS Omni 1 cohort. For the rest of the FHS cohorts, the baseline examination was defined as follows: Original cohort, visit 14; Offspring cohort, visit 4; New offspring spouse cohort, visit 2; Generation 3 cohort, visit 2; and Omni 2 cohort, visit 2. Participants with dementia at baseline or missing both race and ethnicity data were excluded from the present analysis. The sample size and baseline age distribution by cohort is reported in the appendix (p 6).

DRPP and UK Biobank data use at Northwestern University and the University of Minnesota has been approved by their Institutional Review Boards. The individual cohorts have been approved by their local Institutional Review Boards and received written informed consent from participants.

Procedures and outcomes

Data from ARIC, FHS, MESA, and Whitehall II were previously pooled and rigorously harmonised by the DRPP.[13] UK Biobank data were harmonised using the same

approach as all cohorts included in the DRPP consortium. In brief, demographic, clinical, and lifestyle data were re-coded and categorised to yield comparable variable metrics (eg, race and ethnicity categories, and alcohol use quantities and categories) across cohorts that might have collected data using different metrics.[14] We used the psHarmonize package (available on CRAN) to manage harmonisation documentation and to assist with evaluating the harmonisation process.

Demographic variables used in the analysis include self-reported sex, self-reported race and ethnicity (Hispanic, White, Black, Asian, and Other), and education (grade school education or less, at least some high school, and at least some post-secondary school). Race and ethnicity was categorised such that if a participant indicated Hispanic ethnicity, their value for this variable is Hispanic. If a participant did not indicate Hispanic ethnicity, their value for this variable is their self-reported race. Clinical and lifestyle risk factors were measured at the baseline visit and were hypertension (non-elevated blood pressure, elevated blood pressure, hypertension stage 1, and hypertension stage 2), diabetes (no diabetes, prediabetes, and diabetes), obesity (not overweight or obese, overweight, and obese), hypercholesterolaemia (non-elevated cholesterol, elevated, and hypercholesterolaemia), depression (yes or no), alcohol overconsumption (>168 g per week), smoking (current, former, and non-smoker), and physical inactivity (0 metabolic equivalent of a task [MET] minutes per week of moderate or vigorous exercise). Exact definitions for hypertension, diabetes, obesity, hypercholesterolaemia, and depression are in the appendix (p 7). Details on the methodology are provided in the appendix (p 2).

The primary endpoint was dementia, with a cutoff of age 65 years for defining early-onset dementia. Dementia was ascertained according to the protocol of each contributing study (appendix p 8). Clinical assessments were used to identify cases of dementia in FHS and ARIC, and administrative codes were used to identify cases of dementia in all cohorts (appendix pp 9–10). Secondary endpoints were based on changing the age cutoff for early-onset dementia and dementia subtypes. Details on the methodology are provided in the appendix (pp 2–3).

Statistical analysis

Details of the statistical methods are provided in the appendix (pp 3–4). In brief, harmonised data from all cohorts were pooled for a combined analysis and multiple imputation was used to address missing values. The time at risk for dementia began at the baseline visit and continued until dementia diagnosis, death, loss to follow-up, or administrative censoring, whichever occurred first. The time at risk for early-onset dementia analyses was censored at age 65 years, and the time at risk for late-onset dementia analyses began at age 65 years. Crude dementia incidence rates were calculated in this manner for early-onset dementia and late-onset dementia and further stratified by sex. Incidence rates were also calculated per 5-year age group (50–54 years, 55–59 years, 60–64 years, 65–69 years, 70–74 years, 75–79 years, 80–84 years, ≥85 years), and the US Census statistics from 2023 were used to calculate the approximate number of new cases of early-onset dementia in a given year based on these incidence rate estimates.[15]

The association of each exposure (demographic characteristics and clinical and lifestyle risk factors) with early-onset dementia and late-onset dementia, and a test for whether the association differs, was estimated using Cox regression models with time-varying coefficients. Separate Cox models were fitted for each exposure of interest. Cox models used participant age as the timescale. Time-varying coefficient models were estimated that allowed the effect of the exposure of interest to change at age 65 years. From these models we estimated hazard ratios (HRs) for early-onset dementia, late-onset dementia, and the interaction between each of the exposures and time (early onset *vs* late onset). Adjustment for confounding was incorporated via three progressive sets of models: model 1 was unadjusted; model 2 adjusted for cohort, sex, race and ethnicity, and education; and model 3 adjusted for model 2 variables plus all clinical and lifestyle risk factors. Only one variable, the exposure of interest, was modelled with a time interaction within a given model.

In an additional set of analyses, the age cutoff was varied to 60 years, 70 years, 75 years, and 80 years. Dementia subtypes Alzheimer's, vascular, and frontotemporal were also investigated in an additional set of analyses in the UK Biobank cohort. Several sensitivity analyses were conducted (appendix pp 4–5). In brief, they were to exclude UK Biobank, analyse within individual cohorts, implement 3-year and 10-year lag periods, restrict follow-up to 10 years and 15 years, de-couple medication use from risk factor classification, exclude participants who were underweight at baseline, use sex-specific thresholds for alcohol consumption, adjust for family history of dementia, model all risk factor estimates in a combined model, analyse complete-case data, and estimate associations with dementia-free mortality. All analyses were conducted using R (version 4.3.2) and Rstudio (version 2024.04.2+764). A threshold of alpha 0·05 was used for statistical significance.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

42 553 participants from the DRPP cohorts (ARIC, FHS, MESA, and Whitehall II) had no missing dementia status and 502 131 participants from UK Biobank had no missing dementia status, resulting in 544 684 participants eligible for the analysis. Of the five combined cohorts, 236 participants had dementia prevalent at baseline and were excluded. A subsequent six participants were excluded due to missing race and ethnicity data. The final analytic sample was 544 442 participants (figure 1) and the median

Human Genetics Center, Department of Epidemiology, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA (P S de Vries PhD)**; Departments of Psychiatry, Neurology, and Epidemiology, University of California, San Francisco, San Francisco, CA, USA** (Prof K Yaffe MD)**; Imperial College London School of Public Health, London, UK** (Prof A Dehghan MD PhD)**; UK Dementia Research Institute at Imperial College London, London, UK** (Prof A Dehghan MD PhD)

Correspondence to:
Dr Sanaz Sedaghat, Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN 55454, USA
sedaghat@umn.edu

See Online for appendix

For more information on **psHarmonize** see https://cran.r-project.org/web/packages/psHarmonize/index.html

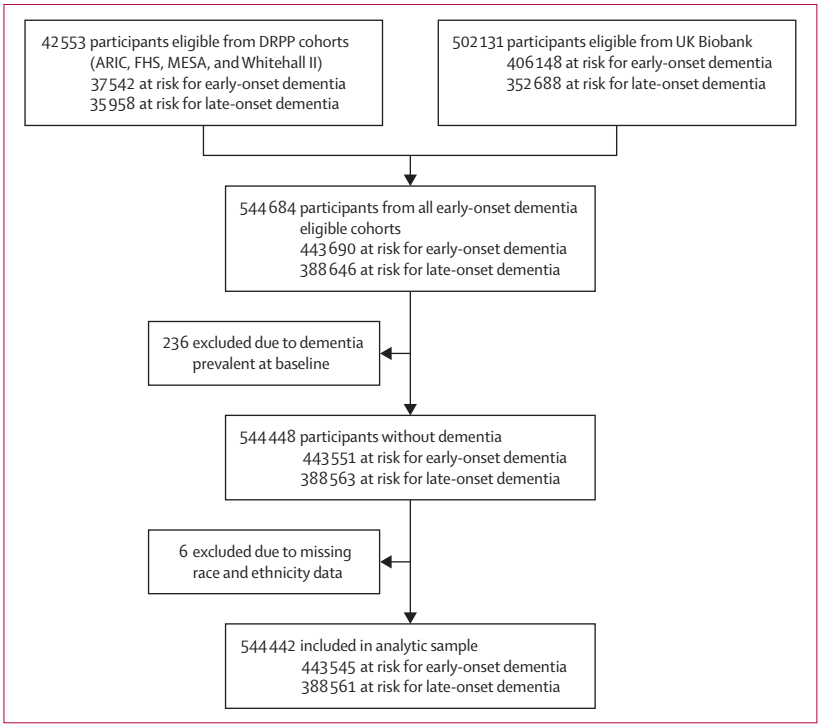

Figure 1: **Study profile**
The same individual could be at risk for both early-onset dementia and late-onset dementia. ARIC=Atherosclerosis Risk in Communities. DRPP=Dementia Risk Prediction Project. FHS=Framingham Heart Study. MESA=Multi-Ethnic Study of Atherosclerosis. The baseline examination was determined according to the time when dementia ascertainment began for each cohort which was as follows: UK Biobank, visit 1; ARIC, visit 1; FHS Original cohort, visit 14; FHS Offspring cohort, visit 4; FHS New offspring spouse cohort, visit 2; FHS Generation 3 cohort, visit 2; FHS Omni 1 cohort, visit 1; FHS Omni 2 cohort, visit 2; MESA, visit 1; Whitehall II, visit 1.

follow-up time was 13·7 years (IQR 12·9–14·4). The mean age at baseline for each cohort was as follows: ARIC, 54·2 years (SD 5·8); FHS, 54·1 years (12·0); MESA, 62·1 years (10·2); UK Biobank, 56·5 years (8·1); and Whitehall II, 45·0 years (6·1). The mean age at baseline in the pooled data was 56·3 years (SD 8·3), and 250 414 (46·0%) participants were male (table). Most participants were White (504 460 [92·7%]). Most participants had at least some post-secondary education (318 097 [58·5%]), and a minority 93 088 [17·1%]) had a grade school education or less. The baseline prevalence of stage 2 hypertension, diabetes, and hypercholesterolaemia was 43·4% (n=236 457), 5·1% (n=27 608), and 42·9% (n=233 652), respectively.

There were 443 545 participants at risk for early-onset dementia (aged younger than 65 years at baseline) and 388 561 participants at risk for late-onset dementia (follow-up time ends after the age of 65 years). The same participant could be part of the at-risk population for both early-onset dementia and late-onset dementia if their follow-up time spanned both age ranges (total of 544 442 participants). There were 807 cases of early-onset dementia and 14 253 cases of late-onset dementia in the sample. The crude incidence rates were 1·97 (95% CI 1·84–2·11) per 10 000 person-years for early-onset dementia and 41·17 (40·50–41·85) per 10 000 person-years for late-onset

	Total sample
Age, years	56·3 (8·3)
Sex	
Female	294 028 (54·0%)
Male	250 414 (46·0%)
Race and ethnicity	
Other	10 504 (1·9%)
Asian	13 003 (2·4%)
Black	14 739 (2·7%)
Hispanic	1736 (0·3%)
White	504 460 (92·7%)
Education	
Grade school education or less	93 088 (17·1%)
At least some high school	132 155 (24·3%)
At least some post-secondary school	318 097 (58·4%)
Missing	1102 (0·2%)
Systolic blood pressure, mm Hg	136·8 (19·1)
Missing	1356 (0·2%)
Hypertension status	
Hypertension stage 2	236 457 (43·4%)
Hypertension stage 1	161 511 (29·7%)
Elevated blood pressure	57 261 (10·5%)
Non-elevated blood pressure	83 523 (15·3%)
Missing	5690 (1·0%)
HbA_{1c} level	
HbA_{1c}, %	5·5 (0·6)
HbA_{1c}, mmol/mol	36·1 (6·8)
Missing	75 206 (13·8%)
Diabetes status	
Diabetes	27 608 (5·1%)
Prediabetes	75 797 (13·9%)
No diabetes	407 802 (74·9%)
Missing	33 235 (6·1%)
BMI, kg/m^2	27·4 (4·8)
Missing	3163 (0·6%)
Obesity status	
Obesity	131 585 (24·2%)
Overweight	227 917 (41·9%)
Not overweight or obese	181 777 (33·4%)
Missing	3163 (0·6%)
LDL cholesterol, mg/dL	137·0 (34·0)
Missing	47 387 (8·7%)
Hypercholesterolaemia status	
Hypercholesterolaemia	233 652 (42·9%)
Elevated	158 376 (29·1%)
Non-elevated cholesterol	109 782 (20·2%)
Missing	42 632 (7·8%)
Depression	
Yes	55 343 (10·2%)
No	424 271 (77·9%)
Missing	64 828 (11·9%)
Alcohol consumption, g/week	48·0 (0·0–112·0)
Missing	18 699 (3·4%)
Alcohol overconsumption	
Yes	72 109 (13·2%)
No	453 634 (83·3%)
Missing	18 699 (3·4%)

(Table continues on next page)

	Total sample
(Continued from previous page)	
Smoking status	
Current smoker	61 590 (11·3%)
Former smoker	186 052 (34·2%)
Non-smoker	293 683 (53·9%)
Missing	3117 (0·6%)
Physical activity as MET, min per week	960·0 (240–2220)
Missing	120 785 (22·2%)
Physical inactivity	
Yes	58 450 (10·7%)
No	365 207 (67·1%)
Missing	120 785 (22·2%)

Data are mean (SD), n (%), or median (IQR). Data were harmonised across the UK Biobank, Atherosclerosis Risk in Communities study, Framingham Heart Study, Multi-Ethnic Study of Atherosclerosis, and Whitehall II study. All risk factors were measured as prevalent at baseline. Race and ethnicity was categorized such that if a participant indicated Hispanic ethnicity, their value for this variable is Hispanic. If a participant did not indicate Hispanic ethnicity, their value for this variable is their self-reported race. HbA_{1c}=glycated haemoglobin A_{1c}. MET=metabolic equivalent of task.

Table: **Baseline patient characteristics (n=544 442)**

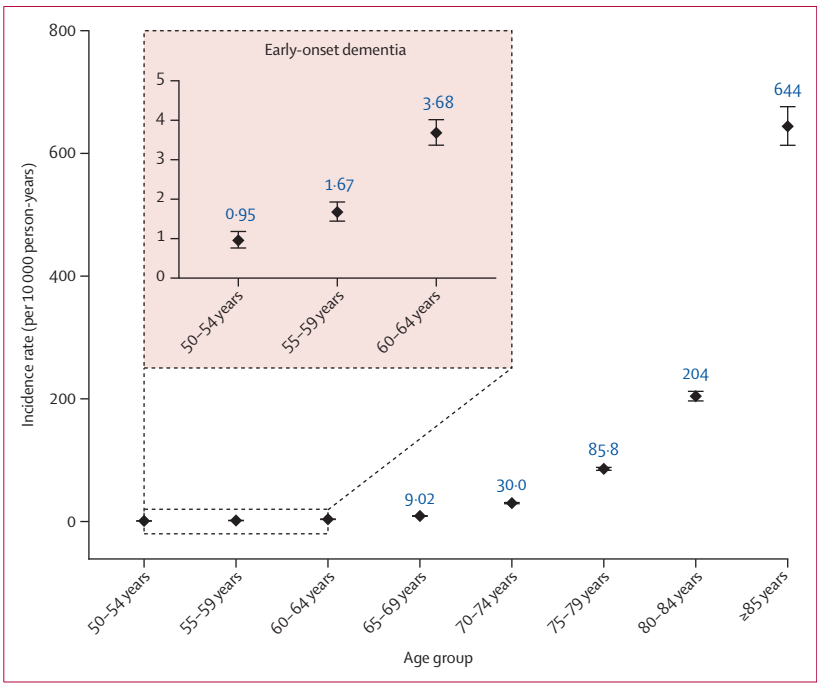

Figure 2: **Dementia incidence rate per 10 000 person-years estimates in 5-year age groups (n=544 442)**
These incidence rates are comparable to the following number of new cases of dementia in the USA in 2023: aged 50–54 years: 1974; aged 55–59 years: 3446; aged 60–64 years: 7820; aged 65–69 years: 17 278; aged 70–74 years: 46 611; aged 75–79 years: 97 755; aged 80–84 years: 142 599; aged 85 years or older: 399 036.

dementia. Stratified by sex, incidence rates were 2·34 (95% CI 2·13–2·57) per 10 000 person-years in male participants and 1·66 (1·49–1·83) per 10 000 person-years in female participants for early-onset dementia, compared with 43·20 (42·19–44·23) per 10 000 person-years in male participants and 39·43 (38·53–40·34) per 10 000 person-years in female participants for late-onset dementia. Crude incidence rates of dementia roughly doubled or tripled every 5 years (figure 2). For example, the incidence rates for the groups 50–54 years, 55–60 years, and 60–64 years were 0·95 (95% CI 0·76–1·18) per 10 000 person-years, 1·67 (1·44–1·93) per 10 000 person-years, and 3·68 (3·37–4·01) per 10 000 person-years, respectively. This finding translates to approximately 13 240 new cases of early-onset dementia each year in persons aged 50–65 years in the USA.

The associations of various demographic, clinical, and lifestyle exposures with early-onset dementia and late-onset dementia from the Cox regression models, as well as a comparison of the strength of those associations are shown (figure 3; appendix pp 11–12). In the fully adjusted model (model 3), female participants had a lower hazard of early-onset dementia compared with males (HR 0·70 [95% CI 0·61–0·80]). Asian participants (HR 1·13 [95% CI 0·75–1·70]), Hispanic participants (1·47 [0·41–5·18]), and those in the remaining racial and ethnic groups (0·94 [0·52–1·70]) did not have notably different hazards of early-onset dementia compared with White participants; whereas the hazard of early-onset dementia in Black participants compared with White participants was higher (1·61 [1·23–2·11]). Lower education was associated with a higher hazard of early-onset dementia, with an HR of 1·44 (95% CI 1·23–1·70) among those with at least some high school education and an HR of 1·99 (1·67–2·38) among those with a grade school education or less, compared with those who obtained at least some post-secondary education.

Clinical and lifestyle risk factors were also associated with a higher hazard of early-onset dementia. Diabetes versus no diabetes (HR 2·45 [95% CI 1·99–3·03]) and depression versus no depression (2·73 [2·34–3·20]) had strong associations with early-onset dementia after adjusting for all demographic, clinical, and lifestyle covariates. Obesity versus not overweight or obese was also associated with higher hazards of early-onset dementia. Compared with their non-elevated ranges, hypertension stage 1 and stage 2 and hypercholesterolaemia exhibited higher, although not statistically significant, HRs for early-onset dementia. From the lifestyle risk factors, current smoking but not former smoking was associated with higher hazards of early-onset dementia, compared with never smoking. Physical inactivity versus being physically active and alcohol overconsumption versus none or moderate consumption were also associated with higher hazards of early-onset dementia (figure 3; appendix pp 11–12).

Several exposures were more strongly associated with early-onset dementia than with late-onset dementia. The protective association between female versus male sex and dementia was more pronounced for early-onset dementia than that for late-onset dementia. Among factors associated with an increased hazard of dementia, the associations for lower versus higher educational attainment, diabetes

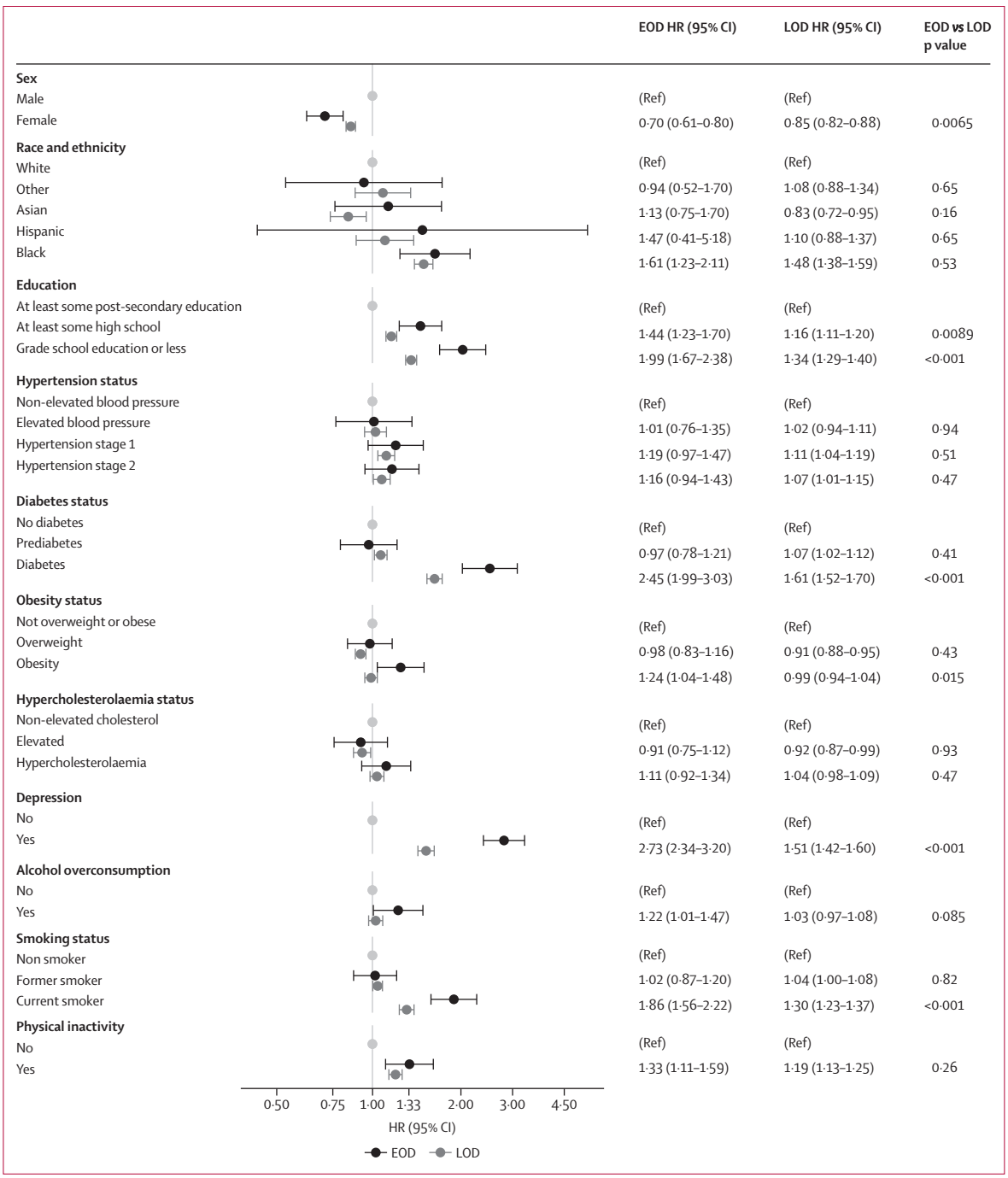

Figure 3: **Differential associations of demographic characteristics and risk factors with early-onset dementia compared with late-onset dementia (n=544 442)**
HR estimates were adjusted for cohort, education, sex, race and ethnicity, and all other risk factors (model 3). Data were harmonised across the UK Biobank, Atherosclerosis Risk in Communities study, Framingham Heart Study, Multi-Ethnic Study of Atherosclerosis, and Whitehall II study. All risk factors were measured as prevalent at baseline. There were 807 early-onset dementia cases from 443 545 participants contributing person-time before age 65 years. There were 14 253 late-onset dementia cases from 388 561 participants contributing person-time after age 65 years. The same individual could contribute person-time to both the early-onset dementia and late-onset dementia time periods. EOD=early-onset dementia. HR=hazard ratio. LOD=late-onset dementia.

versus no diabetes, obesity versus not overweight or obese, depression versus no depression, and current versus never smoking were stronger for early-onset dementia than late-onset dementia.

Excluding UK Biobank, the estimates for depression versus no depression were attenuated for early-onset dementia (HR 1·24 [95% CI 0·70–2·19]) and late-onset dementia (1·10 [0·98–1·23]; appendix pp 13–14). In

analyses for reverse causation, estimates for depression were attenuated in the analysis with a 10-year lag period (appendix pp 15–16). Hypercholesterolaemia versus non-elevated cholesterol was not associated with higher hazards of early-onset dementia or late-onset dementia in the main analysis. However, with medications de-coupled from the definition of clinical risk factors, the use of lipid-lowering medication was associated with higher hazards of both early-onset dementia and late-onset dementia (appendix pp 17). Other sensitivity analyses including repeating the analysis in individual cohorts (appendix pp 18–19), restricting follow-up time (appendix pp 20–21), excluding those who are underweight at baseline (appendix pp 22–23), modelling all variables with time-varying coefficients in a combined model (appendix pp 24–25), using sex-specific thresholds for alcohol consumption (appendix pp 26), adjusting for family history of dementia (appendix pp 27–28), and using complete-case data (appendix pp 29–30), did not essentially change the main findings.

When the age threshold for dementia events was shifted in 5-year increments from 60 years to 80 years, the associations of female versus male sex, lower versus higher educational attainment, diabetes versus no diabetes, depression versus no depression, alcohol overconsumption versus no or moderate consumption, current versus never smoking, and physical inactivity were attenuated (appendix pp 31–33, 38–39). The associations of obesity versus not overweight or obese and hypercholesterolaemia versus non-elevated cholesterol disappeared when the age cutoff shifted to age 70 years or older, and at the same age cutoffs, the association of overweight versus not overweight or obese switched from being null to protective.

Subgroup analyses for sex and age at baseline indicated potential differences in the effects of some factors on the risk of dementia. Results of the sex-stratified analysis (appendix pp 40–41), suggest that hypertension and diabetes might be stronger risk factors for early-onset dementia in female individuals than in male individuals. The analysis stratified by age at baseline accounts for potential differences in association based on life stage. We found that hypertension, diabetes, and obesity might have different effects on early-onset dementia depending on whether the exposure occurs before or after age 57 years (appendix pp 42–43).

In the UK Biobank cohort, frontotemporal dementia accounted for more early-onset dementia cases than late-onset dementia cases (14·2% *vs* 3·1%), whereas Alzheimer's disease accounted for fewer early-onset dementia cases than late-onset dementia cases (51·5% *vs* 61·5%). The proportion of vascular dementia and mixed dementia cases was similar in those who had early-onset dementia compared with late-onset dementia (24·5% *vs* 26·3% and 9·8% *vs* 9·1%, respectively). Comparing the associations for early-onset dementia and late-onset dementia within each dementia subtype, the pattern for each risk factor was generally consistent with the results from all-cause early-onset dementia and late-onset dementia (appendix pp 34–35), with some exceptions. The association of female sex with frontotemporal dementia generally showed a stronger effect for late-onset dementia than early-onset dementia, unlike in the analysis of all-cause dementia, in which it was stronger for early-onset dementia than late-onset dementia. The association of depression with frontotemporal dementia was essentially the same for late-onset dementia and early-onset dementia, unlike in the analysis of all-cause dementia where the association was stronger for early-onset dementia than late-onset dementia. Individual estimates of note include the HR for early-onset vascular dementia, which was particularly large for Asian and Black participants compared with White participants; however, these estimates are based on a small number of cases in each exposure group (n=6). The HR for the association of depression versus no depression with early-onset vascular dementia was also particularly large (5·44 [95% CI 3·52–8·40]). Finally, comparing the associations with early-onset dementia across subtypes, female versus male sex, obesity versus not overweight or obese, depression versus no depression, current versus never smoking, and physical inactivity versus being physically active all had higher HRs for early-onset vascular dementia than early-onset Alzheimer's disease, although this comparison was not tested for statistical significance.

Most risk factor associations with dementia-free mortality were generally similar to their associations with dementia (appendix pp 36–37). Lower education, hypertension stages 1 and 2, prediabetes and diabetes, obesity, depression, alcohol overconsumption, current and former smoking, and physical inactivity were associated with higher hazards of dementia-free mortality both before and after age 65 years. Female sex, Asian versus White race, overweight versus not overweight or obese, and hypercholesterolaemia were protective against dementia-free mortality both before and after age 65 years.

Discussion

In this analysis, we estimated the association of several established late-onset dementia risk factors and demographic characteristics with early-onset dementia, and we compared these associations with that of late-onset dementia. We found that male versus female sex, Black versus White race, lower versus higher educational attainment, diabetes versus no diabetes, depression versus no depression, obesity versus not overweight or obese, current versus never smoking, physical inactivity versus activity, and alcohol overconsumption (>168 g/week) versus moderate or no consumption, were all associated with a greater hazard of early-onset dementia. Furthermore, the associations with early-onset dementia, when present, had stronger effect estimates than the association with late-onset dementia. Smoking, depression, diabetes, and low educational attainment as risk factors were strongly associated with early-onset dementia and have notably larger early-onset dementia HRs than late-onset dementia HRs. The association for hypertension and hypercholesterolaemia with early-onset dementia was not

statistically significant; however, the effect estimates were greater than one. The relationship between these risk factors and dementia might be more complex, such as results from a previous study,[16] in which people with the highest risk of dementia were those with midlife hypertension followed by later-life hypotension. Results from the present sensitivity analysis stratifying by age at baseline support this hypothesis; however, our power is still somewhat low, and these analyses should be interpreted as exploratory. Our results for depression were attenuated when a 10-year lag period was used, which aligns with the possibility that depression might be both a risk factor and an early symptom of dementia.[17,18] The main analysis results for the other risk factors were robust to potential reverse causation, exposure misclassification, and confounding control. As the age threshold for dementia events increased, the most common observation was that the strength of the association weakened. There were no instances in which a risk factor was positively associated with dementia at one age of onset group and negatively associated in the other. The distribution of dementia subtypes aligned with findings in other studies in clinical settings. We found that risk factors were generally more strongly associated with vascular dementia than Alzheimer's disease.[19,20]

Previous studies reported some risk factor associations with early-onset dementia in the UK Biobank cohort.[6,7,9] However, these studies are limited in focusing on only socioeconomic risk factors or relying on dichotomous and clinical risk factors ascertained via administrative records or self-report.[6,7,9] The present analysis included data from UK Biobank and additional studies of participants with diverse racial and ethnic backgrounds. We also assessed clinical measures such as blood pressure and blood glucose to categorise multiple levels of clinical risk factors rather than creating dichotomous variables based on linked administrative records. For example, using both laboratory measurements taken at study visits and self-reported medication use for defining diabetes, we found a strong association between diabetes and early-onset dementia in our analysis. However, previous cohort studies found associations ranging from none to modest.[7–11,16]

There are several limitations to consider when interpreting these findings. Patients with early-onset dementia are younger, so they likely have fewer comorbid risk factors contributing to opportunities for confounding. However, uncontrolled confounding might still be present across all ages despite adjustment efforts. Because individuals must survive dementia-free to be observed in the dementia follow-up period, differential mortality associated with exposures (eg, smoking, diabetes) could lead to survival-related selection, which could attenuate the observed associations. Cancer diagnosis might contribute to survival bias. We included fatal cancer in the competing risk of death; however, because in this cohort non-fatal cancer does not preclude dementia ascertainment, it was therefore not modelled as a competing risk. Exposures were parameterised based on their value at baseline. However, the longitudinal exposure might be of greater aetiological interest, and the baseline value could be considered a biased measure. Dementia has a long preclinical phase, and some early pathological changes might occur before dementia is detected.[21,22] However, results from sensitivity analyses with 3-year and 10-year lag periods to check for reverse causation did not change the overall conclusions. Dementia was ascertained according to each cohort's established protocol, and misclassification could have occurred in each of these settings to varying degrees. Several cohorts relied on administrative records alone for dementia classification, which might lead to underdiagnosis of early-onset dementia cases; however, specificity and positive predictive values are high, resulting in bias likely toward the null.[23] Dementia subtypes were classified via administrative hospitalisation discharge codes, which might have low accuracy. Combining all dementia subtypes together in some cohorts might mask important patterns that are specific to the subtype. The UK Biobank cohort is known to be healthier and more educated than the general population, which could reduce the external validity of the results.[24]

This study also has several strengths. Through pooling several prospective cohorts, we obtained a relatively large number of incident early-onset dementia cases that allowed for higher complexity and dimensionality of models than what is possible from individual cohorts. We formally tested whether associations are different in early-onset dementia than late-onset dementia, we examined how the age cutoff for defining early-onset disease changes the results, and we included a follow-up analysis of dementia subtypes.

In summary, several known modifiable risk factors for late-onset dementia are also risk factors for early-onset dementia, and a number of these factors might be more strongly associated with early-onset dementia than with late-onset dementia. The incidence of dementia in midlife roughly doubles every 5 years, and although early-onset dementia is rare, the effect of receiving this diagnosis in midlife is devastating. These findings highlight the importance of modifiable risk factors in the development of early-onset dementia and contribute to our understanding of which risk factors might be the most important targets for primary prevention efforts.

Contributors

SaS and RM designed the study, with additional methodological contributions from KG, JJS, MM, ALG, DAL, AS, FJW, LJL, OLL, KY, AB, TMH, AD, DS, and PLL. SuS, CLS, JJH, TMH, and AS-M contributed participant data resources. JJS, EAP, and KG curated said data. KG and JJS did the formal analysis. NBA and SaS contributed computing resources. KG and SaS had full access to all of the data and verified its accuracy. KG wrote the original draft and all authors (JJS, MM, EAP, ALG, EMB, DAL, CH, SB, AS, MAI, FJW, SuS, CLS, JJH, LJL, DL, DM, DML-J, FAS, LZ, OLL, SEJ, TMH, VG, AEA, AF, AS-M, BS, JSP, PSdV, KY, AB, AD, RZ, DS, PLL, NBA, RM, SaS) contributed to review and editing. SaS supervised the study and EAP and NBA provided additional project administration support. SaS, NBA, and DML-J obtained funding for the study. All authors approved the final version of the manuscript.

Declaration of interests

PLL declares receiving grants from the National Institutes of Health (NIH) to her institution. AEA declares receiving grants from the NIH to her institution and providing expert testimony on a case related to infectious disease transmission which was not related to this paper in any way. BS declares receiving grants from the NIH which supported his efforts on this paper. CLS declares receiving grants from the NIH and the Texas Alzheimer's Research and Care Consortium to her institution, receiving support for attending meetings and/or travel from the NIH and the University of Rio Grande Valley, and receiving payments for her role on the Framingham Heart Study Executive Committee. DS declares receiving grants from the NIH as a co-investigator on a grant for this project. EMB declares receiving grants from the NIH to her institution and to herself. JSP declares receiving grants from the NIH to his institution. JJS declares receiving grants from the NIH to his institution. NBA declares receiving grants from the NIH. FJW declares receiving grants from The Netherlands Organization for Health Research and Development (ZonMw); Alzheimer's Association; Dutch Heart Foundation; Erasmus Trust Fund; Alzheimer Nederland; and ABOARD partnership funded by the Netherlands Organization for Health Research and Development (ZonMw), Health Holland, and Topsector Life Sciences & Health to his institution; participating on a data and safety monitoring board (DSMB) or advisory board for European Medicines Agency (payment to self) and Health Council of the Netherlands (payment to institution); and serving on the executive committee for ISTAART and The International Society of Vascular Behavioural and Cognitive Disorders, both roles unpaid. TMH declares receiving grants from the NIH. SEJ declares receiving grants from the NIH, receiving funding from the NIH for travel to a meeting, and serving on two NIH DSMBs. PSdV declares receiving grants from the NIH to his institution. RM declares receiving grants from the NIH. JJH declares receiving grants from the NIH, Alzheimer's Association, and William Castella Family to his institution. AF declares receiving grants from the NIH. DAL declares receiving grants from the NIH, receiving consulting fees for work on NIH grants from Northwestern University and Tufts University, and participating on a DSMB for an NIH grant testing the effectiveness of a health services intervention for improving blood pressure control in older adults. KY declares receiving grants from the NIH. DML-J declares receiving grants from the NIH to his institution, participating on an NIH DSMB or advisory board, serving in a leadership or fiduciary role for the American Heart Association (unpaid), and part-time employment by the American Heart Association. All other authors declare no competing interests.

Data sharing

Reasonable de-identified data requests can be submitted to be reviewed by DRPP and individual cohorts to ensure that data can be shared without compromising patient confidentiality or breaching intellectual property restrictions. Participant-level demographic and clinical data might be partially restricted based on previously obtained participant consent. A data dictionary is available on the same webpage. All data access is subject to institutional and research ethics board approvals.

Acknowledgments

This research has been conducted using the UK Biobank Resource under application number 102896. This work uses data provided by patients and collected by the National Health Service as part of their care and support. We would like to thank the staff and study participants from the ARIC, FHS, MESA, and Whitehall II studies, as well as the following data managers and analysts at each cohort for their time, effort and collaboration: Lindsay Clayson, Aurore Fayosse, Lisa Reeves, and David Vu. A full list of participating investigators and institutions for each cohort can be found online (https://aric.cscc.unc.edu/aric9/, https://www.framinghamheartstudy.org/, http://www.mesa-nhlbi.org, and https://www.ucl.ac.uk/population-health-sciences/epidemiology-health-care/research/ucl-research-department-epidemiology-public-health/research/whitehall-ii).

SaS is supported by the US NIH–National Institute of Aging (NIH/NIA) via grant R01AG079108. NBA and the DRPP are supported by the US NIH–National Institute for Neurologic Disorders and Stroke (NIH/NINDS) via grants 1R61NS120245–01/R33NS120245. DAL reports funding support from the NINDS (grant 1RF1NS136499) and the NIA (grant 4R01AG068410). ALG was supported by the NIA (grants R01AG030153 and R01AG070953). TMH was supported by relevant grant funding from the NIA (grants P30AG072947, R01AG054069, R01AG058969, and U01HL096812). OLL receives support from the NIA (grant R01AG20098). AD receives support from the UK Dementia Research Institute (award number UKDRI-5202 and UKDRI-1003) through UK DRI, principally funded by the Medical Research Council. EMB receives support from the NIA (grant K23AG080035). CLS receives support from NIA R01AG059727 and R01AG082360 and NINDS UF1/UH1 NS125513. CLS, JJH, and SuS are partially supported by the South Texas Alzheimer's Disease Research Center (grant P30AG066546). SuS and JJH receive support from The Bill and Rebecca Reed Endowment for Precision Therapies and Palliative Care. JJH is supported by an endowment from the William Castella family as William Castella Distinguished University Chair for Alzheimer's Disease Research, and SuS by an endowment from the Barker Foundation as the Robert R Barker Distinguished University Professor of Neurology, Psychiatry and Cellular and Integrative Physiology. KY is supported by the NIA (grants R01AG063887, R35AG071916, and R01AG091431). Drs FJW and AI are partially supported by the Netherlands Organisation for Health Research and Development (ZonMw project BIRD-NL-10510032120005). BS is supported by NIA (grant 5R21AG081689). This work was funded by the National Heart Lung and Blood Institute (NHLBI; Framingham Heart Study Contracts number N01-HC-25195, number HHSN2682015000001I, and number 75N92019D00031), Boston University School of Medicine, and grants from the NIA (R01AG054076, R01AG049607, U01AG052409, R01AG059421, RF1AG063507, RF1AG066524, and U01AG058589) and the NINDS (R01NS017950). The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the NHLBI under contract numbers 75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, and 75N92022D00005. The ARIC Neurocognitive Study is supported by grants U01HL096812, U01HL096814, U01HL096899, U01HL096902, and U01HL096917 from the NIH (NHLBI, NINDS, NIA, and NIDCD). MESA was supported by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169 from the NHLBI, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences. This paper has been reviewed and approved by the MESA Publications and Presentations Committee. LJL, DL, and DM were supported by the Intramural Research Program, NIA, and NIH. Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

For more information on **submitting requests** see https://drpp.northwestern.edu/research/

References

1 Feng S, Wang T, Su Y, et al. Global burden, risk factors, and projections of early-onset dementia: Insights from the Global Burden of Disease Study 2021. *Ageing Res Rev* 2025; **104:** 102644.
2 Fang M, Hu J, Weiss J, et al. Lifetime risk and projected burden of dementia. *Nat Med* 2025; **31:** 772–76.
3 Rossor MN, Fox NC, Mummery CJ, Schott JM, Warren JD. The diagnosis of young-onset dementia. *Lancet Neurol* 2010; **9:** 793–806.
4 Reitz C, Rogaeva E, Beecham GW. Late-onset vs nonmendelian early-onset Alzheimer disease: a distinction without a difference? *Neurol Genet* 2020; **6:** e512.
5 Mendez MF. Early-onset Alzheimer disease and its variants. *Continuum (Minneap Minn)* 2019; **25:** 34–51.
6 Li R, Li R, Xie J, et al. Associations of socioeconomic status and healthy lifestyle with incident early-onset and late-onset dementia: a prospective cohort study. *Lancet Healthy Longev* 2023; **4:** e693–702.
7 Hendriks S, Ranson JM, Peetoom K, et al. Risk factors for young-onset dementia in the UK Biobank. *JAMA Neurol* 2024; **81:** 134–42.
8 Chun MY, Chae W, Seo SW, et al. Effects of risk factors on the development and mortality of early- and late-onset dementia: an 11-year longitudinal nationwide population-based cohort study in South Korea. *Alzheimers Res Ther* 2024; **16:** 92.

9 Shang X, Zhu Z, Zhang X, et al. Association of a wide range of chronic diseases and apolipoprotein E4 genotype with subsequent risk of dementia in community-dwelling adults: a retrospective cohort study. *EClinicalMedicine* 2022; **45:** 101335.

10 Nordström P, Nordström A, Eriksson M, Wahlund LO, Gustafson Y. Risk factors in late adolescence for young-onset dementia in men: a nationwide cohort study. *JAMA Intern Med* 2013; **173:** 1612–18.

11 Hu Y, Korhonen K, Li P, Bobak M, Martikainen P, Bijlsma MJ. Association between alcohol use disorders and dementia in 262,703 dementia-free Finnish adults: is cardiovascular disease a mediator? *J Gerontol A Biol Sci Med Sci* 2023; **78:** 1045–52.

12 UK Biobank. Our participants. 2025. https://www.ukbiobank.ac.uk/about-our-data/our-participants/ (accessed March 5, 2026).

13 Krefman AE, Stephen J, Carolan P, et al. Cohort profile: Dementia Risk Prediction Project (DRPP). *Int J Epidemiol* 2024; **53:** dyae012.

14 Stephen JJ, Carolan P, Krefman AE, et al. psHarmonize: Facilitating reproducible large-scale pre-statistical data harmonization and documentation in R. *Patterns (N Y)* 2024; **5:** 101003.

15 US Census Bureau, Population Division. Annual estimates of the resident population for selected age groups by sex for the United States: April 1, 2020 to July 1, 2023. In: NC-EST2023-AGESEX. Release date June 2024. https://www2.census.gov/programs-surveys/popest/tables/2020-2023/national/asrh/ (accessed March 9, 2026).

16 Walker KA, Sharrett AR, Wu A, et al. Association of midlife to late-life blood pressure patterns with incident dementia. *JAMA* 2019; **322:** 535–45.

17 Byers AL, Yaffe K. Depression and risk of developing dementia. *Nat Rev Neurol* 2011; **7:** 323–31.

18 Dintica CS, Habes M, Schreiner PJ, Launer LJ, Yaffe K. Trajectories in depressive symptoms and midlife brain health. *Transl Psychiatry* 2024; **14:** 169.

19 Dharmasaroja PA, Assanasen J, Pongpakdee S, et al. Etiology of dementia in Thai patients. *Dement Geriatr Cogn Disord Extra* 2021; **11:** 64–70.

20 Edahiro A, Miyamae F, Taga T, et al. Incidence and distribution of subtypes of early-onset dementia in Japan: a nationwide analysis based on annual performance reports of the Medical Centers for Dementia. *Geriatr Gerontol Int* 2020; **20:** 1050–55.

21 Peters R, Peters J, Booth A, Anstey KJ. Trajectory of blood pressure, body mass index, cholesterol and incident dementia: systematic review. *Br J Psychiatry* 2020; **216:** 16–28.

22 Wagner M, Helmer C, Tzourio C, Berr C, Proust-Lima C, Samieri C. Evaluation of the concurrent trajectories of cardiometabolic risk factors in the 14 years before dementia. *JAMA Psychiatry* 2018; **75:** 1033–42.

23 Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

24 van Alten S, Domingue BW, Faul J, Galama T, Marees AT, Reweighting UK. Reweighting UK Biobank corrects for pervasive selection bias due to volunteering. *Int J Epidemiol* 2024; **53:** dyae054.