European Journal of Epidemiology
https://doi.org/10.1007/s10654-026-01430-1

COHORT PROFILE

EPIC4ND—European Prospective Investigation into Cancer and Nutrition follow-up for neurodegenerative diseases

Christina M. Lill[1,2] · Jan Homann[1] · Olena Ohlei[1] · Karl Smith-Byrne[3] · Vivian Viallon[4] · Jose Maria Huerta[5,6] · Fanny Artaud[7] · Yujia Zhao[8] · Abigail Britten[9] · Laura Deecke[1] · Valerija Dobricic[10] · Olatz Mokoroa[11,12] · Marcela Guevara[13,14,15] · Dafina Petrova[16,17,14] · Miles Trupp[18] · Jara Sabin[19,20] · Mareike Gröninger[19] · Sandra M. Colorado-Yohar[5,6,21] · Sonia Alonso-Martin[22] · Natalia Cabrera-Castro[5,6] · Sergiu Groppa[23] · Oliver Robinson[2,24] · Johnni Hansen[25] · Sture Eriksson[26] · Sabina Sieri[27] · Ana Jimenez[28,29] · Tammy Y. N. Tong[3] · Rudolph Kaaks[19] · Gianluca Severi[7,30] · Ruth Travis[3] · Nick Wareham[9] · Andréa L. Benedet[31] · Henrik Zetterberg[31,32,33,34,35,36] · Andre Franke[37] · Alexis Elbaz[7] · Lars Bertram[10] · Roel Vermeulen[8,38] · Lefkos Middleton[2] · Giovanna Masala[39] · Carlotta Sacerdote[40] · Susan Peters[8] · Verena Katzke[19] · Pietro Ferrari[4] · Marc J. Gunter[24] · Elio Riboli[24]

Received: 29 January 2025 / Accepted: 9 June 2026

Abstract
The 'European Prospective Investigation into Cancer and Nutrition' cohort (EPIC) is a prospective study including ~ 520,000 participants recruited across Europe (1991–2000) with in-depth baseline data on nutritional, lifestyle, medical, and anthropometric variables, and baseline blood samples. Here we introduce EPIC4ND, a case-cohort study within EPIC designed to identify biomarkers predicting a future onset of dementia, Alzheimer's disease (AD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS). EPIC4ND comprises 6415 initially non-diseased participants (aged 35–80 years, mean age at baseline: 54 ± 9, 64% women) including 1899 incident cases with up to 30 years of follow-up and data on at least one omics domain available from pre-disease blood samples. EPIC4ND includes 4604 subcohort members (4441 non-cases and 163 incident cases) and 1811 additional incident cases ascertained from the broader EPIC cohort. Among the incident cases, there are 1190 dementia cases (818 AD), 610 PD cases, and 199 ALS cases. Additionally, 72 prevalent PD cases and 118 incident Parkinsonism cases are available for comparison. Molecular data generated encompass proteomics, genome-wide DNA methylation, and SNP genotyping with 4127 EPIC4ND participants (including 1635 incident cases) having data on all three domains. Smaller studies include data on metals, metabolites, and environmental chemicals, while ongoing efforts focus on ultrasensitive targeted biomarker measurements and small RNA sequencing. Genome-wide association studies and analyses of epidemiological risk factors validate the dataset by confirming many known risk factors. Leveraging these extensive pre-disease multi-layered *omics* data offers a unique opportunity to identify biomarker signatures predicting neurodegenerative diseases and to explore their interplay with epidemiological risk factors.

Keywords Biomarker · Alzheimer · Parkinson · Amyotrophic lateral sclerosis · Neurodegeneration · Multi-omics

The original EPIC cohort

The 'European Prospective Investigation into Cancer and Nutrition' cohort (EPIC) is a multi-centre prospective study including 519,978 participants (70.5% women, mostly aged 35–70 years at baseline) recruited from 23 centers across ten European countries [1, 2]. With approximately 30 years of follow-up, EPIC was originally designed to explore the

Christina M. Lill and Jan Homann are shared first authors.

Marc J. Gunter and Elio Riboli are shared last authors.

Extended author information available on the last page of the article

Published online: 22 July 2026

relationships between diet and cancer, while also having the potential to investigate other chronic diseases. Between 1991 and 2000, the original EPIC cohort recruited participants from the general adult population (generally from 35 to 70 years) across 10 European countries. The sampling frame varied by country and center as detailed e.g. in ref. [2]. Briefly, most centers recruited adults from the general population residing in a specific town or area. However, there were a few exceptions to this recruitment scheme, of which the following are of relevance for EPIC4ND: Parts of the Italian and Spanish cohorts included members of local blood donors, employees of selected enterprises, and/or civil servants; the cohort in Utrecht (The Netherlands) and a subfraction of the cohort in Florence (Italy) included women invited for a local population-based breast cancer screening programme. In Oxford (UK) half of the cohort was recruited among subjects who did not eat meat, i.e., vegans, vegetarians, and fish eaters. In Utrecht, only women were recruited. As a result, a fraction of participants recruited into the EPIC cohort have led healthier lifestyles or have had higher educational attainment than the general population; thus, the cohort is not perfectly generalizable to the general population at mid-life. Eligible individuals were invited via postal mail or in person. Individuals who agreed to participate signed an informed consent agreement. In most centers, participants received a questionnaire on diet and a questionnaire on lifestyle via mail. Most participants completed these questionnaires at home and were then invited to a study centre for an examination. This included collection of the two completed questionnaires, anthropometry and measurement of blood pressure. Blood samples were collected and separated into citrate plasma, serum, buffy coats, and erythrocytes using identical protocols and equipment, and then aliquoted for long-term storage in liquid nitrogen (− 196 °C) at a central biobank located at the International Agency for Research on Cancer (IARC)/World Health Organization in Lyon, France [1, 2]. A computer-assisted 24-h dietary recall was recorded in a representatitive subsample (n = 36,900) for calibration purposes [3]. Not all of the centres maintained records of all individuals invited to participate. Based on centers relevant for EPIC4ND with corresponding data available, the response rate ranged between 60% and 35% (Spanish centers: 55–60% [4], Cambridge/Norfolk: 40% [5], Heidelberg: 38% [6], Utrecht: 35% [7]). Follow-up aimed at identifying incident cancer and other diseases has been based on a combination of center-specific methods including linkage to health insurance records, cancer and pathology records, mortality records, and, in a subset of centers, also on active folllow-up. Since the initial enrollment, several working groups have conducted case ascertainments and research projects to investigate the role of lifestyle and dietary factors on a range of outcomes beyond cancer, such as diabetes [8], cardiovascular disease [9], inflammatory bowel disease [10], and neurodegenerative diseases [11–19]. This also offers the possibility to investigate multimorbidity as well as the intersections of risk factors and pathways across chronic diseases.

Rationale for the new focus and new data collection

Neurodegenerative diseases (ND) such as Alzheimer's disease (AD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS) are becoming increasingly prevalent, particularly in aging, industrialized populations, leading to increasing costs for the public health systems [20–22]. For example, the number of people diagnosed with dementia is projected to rise from 55 million in 2019 to 139 million by 2050 globally. As a result, the costs associated with dementia are expected to drastically increase, too, from 1.3 trillion US dollars annually in 2019 to 2.8 trillion US dollars by 2030 (https://www.alzint.org/u/World-Alzheimer-Report-2023.pdf). Given these trends, the ultimate goal of translational research in the ND field is to develop strategies for preventing these devastating and currently incurable conditions. In this context, identifying individuals at risk in pre-disease or early disease stages may be important for improving the success of clinical treatment trials (e.g., ref. [23]). Pathophysiological mechanisms underlying AD and PD begin as early as 20 years or more before diagnosis (while the prodromal period for ALS is largely unknown). Accordingly, the identification of robust and easily accessible predictive blood biomarkers suitable for screening strategies relies on the availability of large prospective cohorts with baseline blood samples and long-term follow-up (i.e., > 10 years). This represents a major challenge, particularly for diseases with lower incidence rates such as PD and ALS. Consequently, relative to diagnostic biomarker studies, few prospective population-based studies with long-term follow-up have explored predictive molecular biomarkers (see below). Candidate blood biomarkers such as p-tau217, Aβ 42/40, GFAP, and neurofilament light chain (NfL) possess predictive potential in the path towards clinical AD, especially in settings when other biological AD readouts are already positive (e.g., based on Aβ positivity in PET scanning or CSF), when focusing on conversion to AD over short follow-up intervals of a few years and/or in select individuals at higher risk of AD (e.g., [24–26]). However, their long-term performance in the general population remains less explored, and current evidence suggests that these biomarkers must be combined with other predictors for sufficient predictive performance [27–31]. For PD, possibly one of the most promising, and potentially also predictive, candidate biomarkers

is abnormally aggregated α-synuclein, which has reportedly been detected in seed-amplification assays in different tissues and biofluids including blood [32]. However, epidemiological studies on prediagnostic blood samples are currently scarse (e.g., ref. [33]; reviewed by ref. [34]). Lastly, for ALS, the hitherto most promising biomarker appears to be NfL (systematically reviewed by ref. [35]); however, data from prospective epidemiological cohorts are limited [29, 36]. Furthermore, NfL lacks disease specificity and is considered as a "general" biomarker for neurodegeneration [29]. Large-scale prospective omics studies, particularly those focusing on proteomics, have only recently begun to emerge, identifying promising additional candidates in ND including AD and dementia [37–40] as well as PD [41]. To address this still largely understudied field, we established the EPIC4ND case-cohort with the aim to develop prediction models for dementia (particularly AD), PD and ALS based on blood-based biomarkers in combination with pre-diagnostic lifestyle factors. The newly established EPIC4ND case-cohort will leverage the existing unique EPIC cohort by generating and analyzing multidimensional molecular biomarkers derived from blood samples taken at baseline, i.e., at early pre-disease stages. This case-cohort design represents a cost-effective and powerful alternative to analyzing the entire EPIC cohort, for which multi-omic profiling is currently unfeasible due to financial and logistical constraints. Furthermore, compared to nested case–control studies, the case-cohort design offers flexibility and efficiency by enabling the simultaneous investigation of multiple outcomes within a single randomly selected subcohort of non-cases, maximizing the number of available non-cases and minimizing the number of biomarker measurements required saving sample resources and costs [42].

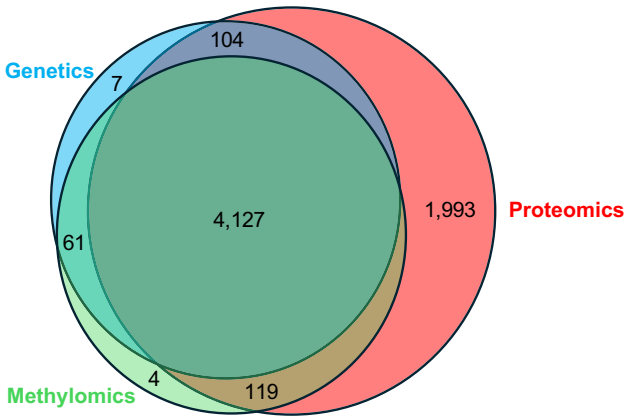

Fig. 1 Venn diagram illustrating the omics domains analyzed in the EPIC4ND case-cohort. Legend. This figure depicts the number of samples from the EPIC4ND case-cohort with measurements available for at least one *omics* domain after quality control. Notably, small RNA-Seq data for 3992 samples and data on established neurodegenerative disease biomarkers for 1700 samples of the core EPIC4ND dataset (i.e., those with data from all three *omics* domains) are currently being generated

New areas of research

To identify molecular signatures and biomarkers predictive of a future onset of AD/dementia, PD, and ALS, several layers of molecular *omics* data were recently generated in baseline blood samples of up to 6,415 EPIC participants using high-throughput methodologies (Fig. 1). This includes proteomic data (SomaScan) generated from plasma samples and genomic and epigenomic (i.e., methylomic) data from blood-derived DNA samples (see below). In addition, the generation of transcriptomic (small RNAs) data and the targeted protein biomarker data for amyloid and tau pathology and neurodegeneration (ultrasensitive immunoassays) is currently underway. The new areas of research will focus on: i) identifying novel individual molecular biomarkers for a future onset of AD, PD, and ALS, ii) discovering molecular biomarker signatures by combining multiple markers within and across molecular domains, iii) investigating biomarkers and pathways shared across the investigated ND and in comparison to other chronic diseases, and iv) integrating these signatures with established demographic and lifestyle risk factors for predictive modelling. The EPIC4ND case-cohort will serve as a valuable resource for addressing a range of additional related questions by exploiting and integrating the available high-dimensional data in the framework of the EPIC study: This includes assessing interactions among markers within the same and across different molecular domains, as well as their interplay with demographic, medical, and lifestyle factors. It also involves exploring the role of biological aging through the construction of molecular clocks and uncovering pathophysiological mechanisms underlying our biomarker observations. Additionally, this multi-*omics* resource can be utilized for research on other quantitative phenotypes measured within the corresponding subcohort in EPIC.

Description of the case-cohort and case ascertainment

The EPIC4ND case-cohort comprises a total of 6415 participants aged 35–80 years at baseline who were recruited as part of the EPIC study in twelve centers across five countries. Quality-controlled data from at least one molecular domain generated from baseline blood samples is available for each EPIC4ND participant. According to the case-cohort design, the 6415 participants comprised 4604 subcohort members (thereof 4441 non-cases from the subcohort that

did not have incident dementia, PD, and ALS, as well as 163 incident cases) and 1811 additional incident cases ascertained from the broader EPIC cohort. Subcohort members were randomly drawn among all participants 35–80 years at baseline from within the existing EPIC-InterAct subcohort [8]. Varying numbers of centers contributed per disease entity (Table 1, Supplementary Tables 1–4): Since ALS was passively ascertained based on linkage to readily available mortality records in EPIC, the number of particpating centers was large (n = 12; Spain: Murcia, Andalucía/Granada, Gipuzkoa/San Sebastian, Navarra; Italy: Turin, Florence, Varese; United Kingdom: Cambridge/Norfolk, Oxford; Netherlands: Utrecht, Bilthoven; Germany: Heidelberg). Participation of some centers from the original EPIC cohort in the centralized biomarker measurement projects was precluded by legal or data protection restrictions. Nearly all centers participating in EPIC4ND were also able to actively link to and validate PD cases (n = 10, except Andalucía/Granada and Oxford). For the dementia/AD case-cohort, the ascertainment of cases required substantial local resources and was feasible only in the four Spanish centers. Regarding the overall sample size, the number of cases was determined by the number of total incident cases with available plasma samples identified in the screened centers, i.e., 1190 dementia cases, thereof 818 AD, 610 PD cases, and 199 ALS cases (Table 1, Supplementary Tables 1–4). The subcohort as reference group was intentionally designed to be substantially larger than the corresponding numbers of cases, which was achieved for each disease studied, and its final size was primarily determined by the amount of available funding. Notwithstanding, to our knowledge EPIC4ND currently represents the largest methylomic and small RNA transcriptomic study on incident neurodegenerative disease and, for incident PD and ALS, also the largest SomaScan-based proteomic study. Notably, 193 eligible incident ND cases who did not have biomaterial available were excluded from the EPIC4ND case-cohort. However, data from these latter individuals can and have already been included in epidemiological, non-molecular studies investigating the entire EPIC cohort (see examples below).

The dementia/AD, PD, and ALS case ascertainments have been described previously [11, 13, 43, 44]. Details of last ascertainment years per center are provided in Supplementary Tables 1–4. Potential PD and Parkinsonism cases were identified using a combination of multiple health record sources as available in the centers: These included self-reported PD diagnoses and symptoms gathered from baseline or follow-up questionnaires, record linkage with primary care records, hospital discharge databases, National Health System databases (Italy), mortality registries using relevant disease codes (ICD-9: 332, 332.0, 332.1, ICD-10: G20, and ICPC: N87), and prescription databases, where use of antiparkinsonian medications (ATC/DDD codes N04, N04A, N04B; Italian drug prescription database: 038) was ascertained. This linkage approach intended to maximize sensitivity as it is highly unlikely that a PD patient will never receive a relevant code in any health database. Following the initial linkage, the validation phase increased specificity: Records on potential cases underwent systematic review by expert neurologists involving scrutiny of medical records such as general practitioner notes, hospital documents, specialist reports, and drug histories. A standardized template was filled in, which allowed for extraction of clinical data, included a checklist on the UK Brain Bank Criteria [45], an assessment of the quantity and quality of the clinical data available and of the confidence of the expert neurologist in the clinical diagnosis leading to a label on the certainty of the final clinical diagnosis ("definite", "very likely", "probable", "possible") [13]. Subsequently, a total of 66% of the identified potential PD cases which underwent medical record review were confirmed to have classical PD in the validation phase [13]. Updates to the PD ascertainment have been performed by five EPIC centers in Germany (Heidelberg), Italy (Florence and Turin), and the Netherlands (Bilthoven and Utrecht) contributing 239 participants as additional PD cases. For comparison purposes, 72 prevalent PD cases and 118 cases with Parkinsonism other than PD, which are not considered to be part of the main EPIC4ND dataset, have also been included in the molecular profiling (Supplementary Table 5).

The dementia/AD case ascertainment was performed following a similar approach [43, 44]: Potential dementia cases were identified by linking data from all available health databases, including primary care databases, hospital discharge, mortality and drug prescription registries using relevant ICD-9 (290, 331), ICD-10 (F00-F03, G30), ICPC2 (P20, P70, N29, N99), and ATC (N06DA02, N06DA03, N06DA04, N06DX01) codes. Expert neurologists then reviewed all available medical records including outpatient and hospital reports, prescriptions, and diagnostic tests to confirm the diagnosis based on a standardized algorithm. A case was validated as incident dementia if a clinical diagnosis was documented or, in a minority of cases with absence of medical reports, if relevant codes or antidementia drug use were present. Extracted data included clinical data (DSM IV criteria, clinical features and subtypes of dementia, incidence date and neuropsychological assessments), functional status, results of the diagnostic tests, use of antidementia drugs, source, quality and amount of information, and the clinical judgement of the reviewer, who finally established the diagnosis of dementia, its subtype when clinical information was sufficient, and the degree of quality of the data (high, medium, low) and certainty of the diagnosis (probable and possible). Note that exact study-specific

Table 1 Selected baseline characteristics of the EPIC4ND case-cohort

Characteristic	AD		Dementia		PD		ALS	
	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases
N (% women)	818 (69)	1818 (62)	1190 (67)	1785 (62)	610 (46)	4010 (65)	199 (57)	4585 (66)
Median AAE (IQR)	59 (55-63)	49 (43-56)	59 (55-63)	49 (43-55)	59 (54-64)	52 (46-59)	58 (53-65)	52 (46-59)
Median AAD (IQR)	78 (74-81)		78 (73-81)		71 (65-75)		73 (67-77)	
Median AAC (IQR)		71 (65-77)		71 (65-77)		69 (63-76)		70 (64-77)
Median follow-up time (IQR)	19 (16-22)	22 (21-23)	19 (16-21)	22 (21-24)	11 (7-15)	16 (15-18)	14 (9-18)	18 (16-20)
High/medium certainty, n (%)	801 (98)		1117 (94)		352 (58)		n.a.	
Education level, n (%)								
No education	495 (61)	639 (35)	722 (61)	616 (35)	52 (9)	500 (12)	15 (8)	692 (15)
Primary school	229 (28)	705 (39)	330 (28)	699 (39)	220 (36)	1505 (38)	68 (34)	1618 (35)
Technical/professional school	25 (3)	140 (8)	38 (3)	139 (8)	138 (23)	722 (18)	47 (24)	782 (17)
Secondary school	32 (4)	107 (6)	38 (3)	107 (6)	72 (12)	617 (15)	20 (10)	654 (14)
Longer education	31 (4)	214 (12)	49 (4)	213 (12)	100 (16)	566 (14)	31 (16)	683 (15)
Missing or not specified	6 (1)	13 (1)	13 (1)	11 (1)	28 (5)	100 (2)	18 (9)	156 (3)
Smoking status, n (%)								
Never	622 (76)	1007 (55)	865 (73)	984 (55)	303 (50)	1902 (47)	89 (45)	2268 (49)
Former	88 (11)	324 (18)	149 (13)	320 (18)	203 (33)	1038 (26)	69 (35)	1161 (25)
Current	108 (13)	487 (27)	175 (15)	481 (27)	92 (15)	998 (25)	33 (17)	1084 (24)
Missing	0 (0)	0 (0)	1 (0)	0 (0)	12 (2)	72 (2)	8 (4)	72 (2)
Coffee consumption (ml/d), n (%)								
0	146 (18)	202 (11)	214 (18)	196 (11)	53 (9)	297 (7)	22 (11)	360 (8)
0.1-199.9	570 (70)	1267 (70)	827 (69)	1243 (70)	291 (48)	2122 (53)	77 (39)	2470 (54)
200-399.9	61 (7)	265 (15)	95 (8)	262 (15)	78 (13)	548 (14)	27 (14)	596 (13)
400-799.9	16 (2)	43 (2)	19 (2)	43 (2)	137 (22)	720 (18)	48 (24)	792 (17)
>800	1 (0)	4 (0)	1 (0)	4 (0)	35 (6)	220 (5)	21 (11)	253 (6)
Missing	24 (3)	37 (2)	34 (3)	37 (2)	16 (3)	103 (3)	4 (2)	114 (2)
Alcohol drinking pattern (lifetime), n (%)								
Never	249 (30)	347 (19)	343 (29)	343 (19)	33 (5)	406 (10)	12 (6)	549 (12)
Former light	112 (14)	216 (12)	162 (14)	212 (12)	59 (10)	311 (8)	15 (8)	367 (8)
Former heavy	9 (1)	22 (1)	18 (2)	21 (1)	3 (0)	30 (1)	2 (1)	31 (1)
Light	90 (11)	193 (11)	120 (10)	188 (11)	67 (11)	424 (11)	25 (13)	541 (12)
Never heavy	240 (29)	702 (39)	358 (30)	690 (39)	349 (57)	1951 (49)	105 (53)	2171 (47)
Periodically heavy	91 (11)	270 (15)	149 (13)	263 (15)	66 (11)	462 (12)	17 (9)	496 (11)
Always heavy	15 (2)	58 (3)	28 (2)	58 (3)	5 (1)	70 (2)	4 (2)	72 (2)
Missing	12 (1)	10 (1)	12 (1)	10 (1)	28 (5)	356 (9)	19 (10)	358 (8)
Physical activity, n (%)								
Inactive	397 (49)	675 (37)	551 (46)	659 (37)	163 (27)	1012 (25)	51 (26)	1248 (27)
Moderately inactive	241 (29)	595 (33)	378 (32)	582 (33)	210 (34)	1307 (33)	63 (32)	1513 (33)
Moderately active	107 (13)	325 (18)	163 (14)	321 (18)	120 (20)	803 (20)	41 (21)	884 (19)
Active	73 (9)	223 (12)	98 (8)	223 (12)	99 (16)	776 (19)	37 (19)	826 (18)
Missing	0 (0)	0 (0)	0 (0)	0 (0)	18 (3)	112 (3)	7 (4)	114 (2)
BMI (kg/m2), n (%)								
<18.5	1 (0)	1 (0)	1 (0)	1 (0)	7 (1)	22 (1)	3 (2)	27 (1)
18.5-24.9	79 (10)	411 (23)	127 (11)	405 (23)	210 (34)	1596 (40)	96 (48)	1765 (38)
25-29.9	369 (45)	872 (48)	534 (45)	857 (48)	277 (45)	1662 (41)	70 (35)	1910 (42)
>30	369 (45)	534 (29)	528 (44)	522 (29)	116 (19)	730 (18)	30 (15)	883 (19)
Diabetes, n (%)								
Yes	89 (11)	101 (6)	153 (13)	95 (5)	34 (6)	135 (3)	6 (3)	167 (4)
No	729 (89)	1714 (94)	1037 (87)	1687 (95)	424 (70)	3258 (81)	122 (61)	3796 (83)
Missing	0 (0)	3 (0)	0 (0)	3 (0)	152 (25)	617 (15)	71 (36)	622 (14)
Hypertension, n (%)								
Yes	260 (32)	347 (19)	386 (32)	338 (19)	180 (30)	738 (18)	39 (20)	856 (19)
No	557 (68)	1470 (81)	802 (67)	1446 (81)	267 (44)	2228 (56)	83 (42)	2688 (59)
Missing	1 (0)	1 (0)	2 (0)	1 (0)	163 (27)	1044 (26)	77 (39)	1041 (23)
Hyperlipidaemia, n (%)								
Yes	235 (29)	382 (21)	348 (29)	371 (21)	152 (25)	694 (17)	36 (18)	798 (17)
No	577 (71)	1423 (78)	835 (70)	1402 (79)	308 (50)	2682 (67)	95 (48)	3145 (69)
Missing	6 (1)	13 (1)	7 (1)	12 (1)	150 (25)	634 (16)	68 (34)	642 (14)
Cardiovascular problem, n (%)								
Yes	268 (33)	359 (20)	397 (33)	350 (20)	195 (32)	788 (20)	42 (21)	909 (20)
No	548 (67)	1455 (80)	790 (66)	1431 (80)	381 (62)	2730 (68)	134 (67)	3183 (69)
Missing	2 (0)	4 (0)	3 (0)	4 (0)	34 (6)	492 (12)	23 (12)	493 (11)

Legend. This table presents selected baseline characteristics of the EPIC4ND case-cohort. Unless otherwise specified, all variables reflect the status at the time of recruitment. N = number, IQR = interquartile range, AAE = age at examination (recruitment), AAD = age at diagnosis, AAC = age at censoring, BMI = body mass index.

sensitivities and specificities of our case ascertainment approach cannot be provided given the lack of an independent "gold standard" applied to all participants of our study. However, given that our data linkages in EPIC4PD and EPIC4AD relied on a comprehensive approach searching for all relevant related codes in all available health databases as described above we expect a comparatively high sensitivity of our initial case ascertainment. Furthermore, this step was followed by validation by expert neurologists, which improves the specificity of our approach. Theremore, we expect the sensitivities and specificities of our case ascertainment to be at the upper end of published estimates for data linkage-based epidemiological studies (for PD, see e.g. systematic review of relevant studies by ref. [46], for an overview of relevant studies in dementia/AD, see ref. [43]). Since the initial ascertainment, updates in selected centers yielded 435 additional incident dementia cases.

Finally, due to the fatal character of the disease and the median short survival in ALS of 2–3 years in European cohorts [47], case ascertainment relied solely on linkage to mortality records, identifying cases using ICD-10 code G12.2 as an immediate, antecedent, or underlying cause of death. In all published EPIC studies for ALS [11, 12, 14, 15], age at death was used as a proxy for age at diagnosis. The EPIC cohort's centralized and complete mortality record system ensured thorough case identification, despite the common unavailability of true diagnosis age from these records. In this context, we addressed concerns about possibly including prevalent cases by requesting age at diagnosis from four participating centers that linked to the earliest available G12.2 code in their health records (excluding mortality records). Among the 37 individuals from these four centers, only one was found to be a prevalent case. These additional data may be used in sensitivity analyses in biomarker studies in EPIC. In this context, analyses particularly for ALS, but also for dementia/AD and PD may include implementing a "wash-out" period of a few years between study recruitment and inclusion in the analysis. This approach helps to exclude cases diagnosed shortly after enrollment, thereby minimizing the inclusion of prevalent cases. Since the follow-up in EPIC is largely based on passive follow-up using a comprehensive linkage strategy, the attrition rate is neligible due to few participants withdrawing from the study or moving out of the country. Since the original publications [11, 12, 14, 15], the linkage to mortality records was updated centrally at IARC and/or locally for selected centers.

In addition to the EPIC4ND case-cohort, a French PD and ALS case-cohort (257 cases [238 PD, 19 ALS] and 281 non-cases) within the French E3N cohort [48, 49] and a nested AD case–control dataset (n = 352 incident cases, n = 352 controls, with 142 additional longitudinal samples from cases) from Umeå, Sweden [50], are available as an extended molecular data resource within the EPIC4ND framework. Genomic and methylomic data have been generated for the French dataset and will be generated for the Swedish dataset using the same platforms and standardized data processing protocols as for the core EPIC4ND cohort. The generation of proteomic data in both datasets is planned as a next step.

Omics data measurements

In the EPIC4ND case-cohort, proteomic data were generated from baseline plasma samples, while genome-wide DNA methylation and SNP genotyping data were generated from DNA extracted from buffy coats. The generation of small RNA sequencing data and targeted biomarker measurements from plasma samples is underway. In subsets of the nested PD and ALS case–control datasets, mass spectrometry has been performed on erythrocytes to also measure metals (PD and ALS) [15, 16] and on plasma to measure endogenous/biological metabolites and environmental chemicals (PD) [17, 19].

Specific to EPIC4ND, proteomic data were generated from plasma samples using the affinity-based SomaScan v4.1 assay (SomaLogic, Inc.) that measures > 6000 proteins based on ~ 7000 aptamers. After quality control (QC), a total of 7285 aptamers (6381 proteins) across 1932 ND cases and 4411 non-cases are now available for statistical analysis. Methylomic data were generated from DNA extracted from buffy coats using the Infinium MethylationEPIC (v2) array (Illumina, Inc.) that contains ~ 930 k CpG sites. Following QC, 889,259 of these CpG sites are available for analysis across 1730 ND cases and 2581 non-cases. Genome-wide SNP genotyping data were generated using the Global Screening Array v1 (Illumina), which includes custom content for ~ 40 k mostly infrequent genetic variants located in genetic loci relevant for ND and movement disorders. Following QC and genotype imputation using established protocols [51], the final dataset comprises 7,381,754 SNPs in 1706 ND cases and 2593 non-cases available for statistical analysis. The "multi-*omics*" core dataset comprises 962 dementia (thereof 650 AD), 523 PD, and 170 ALS cases with data available on all three molecular domains. Due to differing numbers of EPIC centers per disease entity, the numbers of non-cases in the core dataset are 972 for the AD-specific EPIC4ND case-cohort (EPIC4AD), 2227 for EPIC4PD, and 2587 for EPIC4ALS (Supplementary Tables 1–4).

Furthermore, small RNA sequencing in plasma samples of 3992 EPIC4ND participants (small RNA libraries prepared from the NextFlex Small-RNA-Seq kit (PerkinElmer, Waltham, Massachusetts), followed by sequencing on a

HiSeq 4000 instrument (Illumina, San Diego, California) with 2 × 50 bp reads; data analysis will largely follow the protocol described in ref. [52]) and targeted biomarker measurements in plasma samples of 1700 EPIC4ND participants (measuring biomarkers using the CNS NULISA assay, Alamar, Inc.) of the core dataset are currently being generated. In addition, in two overlapping nested case–control studies for PD and ALS (PD: n = 362 cases, n = 362 matched controls; ALS: n = 107 cases, n = 319 matched controls), data on metals in erythrocyte samples were generated using inductively coupled plasma mass spectrometry [15, 16]. Furthermore, for the nested case–control study in PD, metabolomics in plasma were generated using using both liquid chromatography- high resolution mass spectrometry (LC-HRMS) [17, 19] and gas chromatography-high resolution mass spectrometry (GC-HRMS). LC-HRMS is effective for measuring endogenous/biological metabolites, while both LC-HRMS and GC-HRMS are capable of detecting internal markers for environmental chemicals. Most samples from these nested case–control studies also have corresponding genomic and methylomic data available, as described above. Finally, a large-scale whole exome/genome sequencing project, involving the sequencing of approximately 200,000 EPIC samples, is being planned as a future initiative. This project will also encompass EPIC4ND and the French dataset.

Data analysis in the context of a case-cohort

While case-cohort designs offer substantial economic efficiencies, they require some specific analytical considerations. This is mostly due to the oversampling of cases from outside the subcohort in the dataset. To ensure statistical metrics reflect the original cohort, several methods were proposed in the literature to account for this situation, depending on the underlying analysis: For instance, in Cox regression analyses, Prentice's weighting can be applied where incident cases outside the subcohort only enter the risk set very shortly before the time of their event while all subcohort members (cases and non-cases) constantly remain in the risk set until event/right censoring [53]. Barlow's weighting additionally utilizes the inverse of the sampling fraction as weights for incident cases and non-cases of the subcohort [54]. For an overview and comparison of these and related methods, see ref. [42, 55, 56]. Similarly, the oversampling of cases also needs to be considered in re-classification methods such as the Net Reclassification Index (NRI), since reclassification indices are based on absolute risk differences. As a solution, either only the subcohort may be used, but this may lead to imprecision as all the oversampled cases would be ignored. Alternatively, weighting can be applied by the Horvitz-Thompson estimator weighting each participant (cases and non-cases within or outside the subcohort) by the inverse of their selection probability from the original cohort [57].

Key findings

The EPIC4ND case-cohort comprises 1974 incident ND cases (1190 dementia thereof 818 AD, 610 PD, 199 ALS) and 4441 non-cases across 12 centers from five European countries. The mean age at baseline was 54 (range 35–80) years for all EPIC4ND participants. The mean age at diagnosis for the incident cases was 77 (± 7 standard deviation [SD]) years for dementia and 77 ± 6 years for AD, 70 ± 8 years for PD. Mean age at death for ALS (used as a proxy for age at diagnosis, see above) was 72 ± 9 years. The mean time to diagnosis was 19 ± 4 years for AD, 11 ± 5 years for PD, and 13 ± 6 years for ALS. The mean follow-up time for non-cases (defined as time until the last record of absence of the specific disease, usually at the time of last case ascertainment), was 22 ± 4 years in EPIC4AD, 16 ± 4 years in EPIC4PD, and 18 ± 5 years in EPIC4ALS. In EPIC4AD, there were more women among cases than among non-cases. Future AD and other dementia cases were more likely to consume less coffee than average at recruitment, to be overweight, and have diabetes. Additionally, they tended to have lower educational attainment, be never smokers at recruitment, be physically inactive, and to be less likely to drink alcohol lightly compared to never drinkers but more likely to drink heavily. In EPIC4PD, cases comprised more men than non-cases. Future PD cases were less likely to smoke at recruitment and were more likely to be former light drinkers compared to never drinkers. Finally, in EPIC4ALS, there were more men among cases than among non-cases. Future ALS cases were less likely to be overweight (Table 1, Fig. 2, Supplementary Table 6). These results are largely in line with previous EPIC cohort studies on dementia and AD [44], PD (e.g., ref. [58]) and ALS (e.g., ref. [12]) where instead of the case-cohort, the complete cohort of all participating centers was included as reference group. Furthermore, the positive associations of AD risk with obesity and diabetes, and heavy alcohol consumption, and inverse association with education levels are in line with the major risk factors described in independent studies [59] while the inverse association with smoking is unusual. In contrast, the inverse association of PD with smoking is well established based on numerous independent studies [60]. This highlights that our smaller case-cohort design, optimized for molecular analyses, is not only representative of the larger study population but also accurately reflects many of the established risk factors in the source population (Fig. 2).

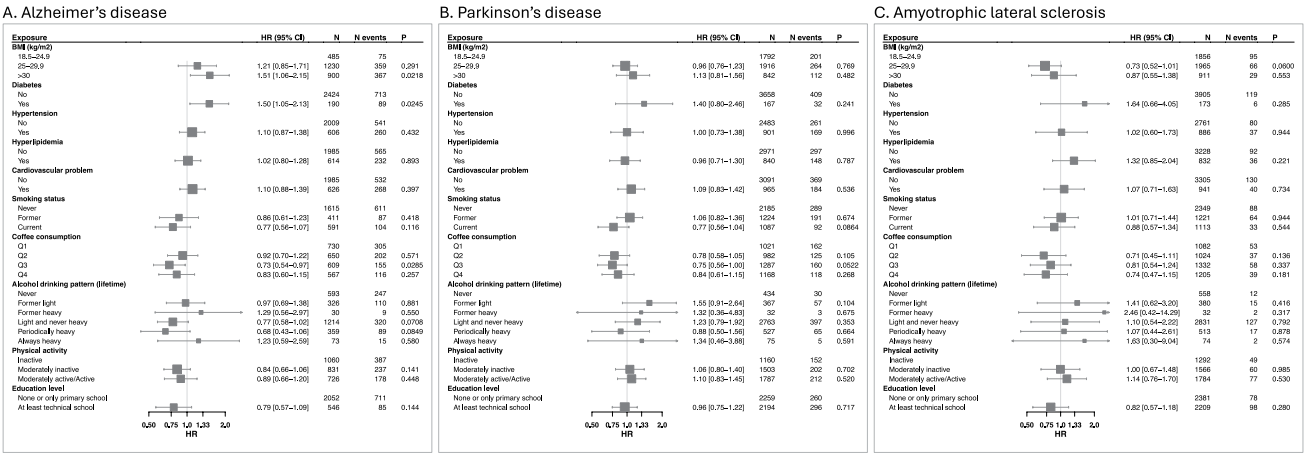

Fig. 2 Forest plots of association analyses of baseline lifestyle and medical factors with instantaneous risk for Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis in the EPIC4ND casecohort. Legend. These forest plots depict the association results of Prentice-weighted Cox proportional hazard regression analyses (stratified for age, sex, and center) on Alzheimer's disease (A), Parkinson's disease (B) and amyotrophic lateral sclerosis (C). HR=hazard ratio, CI=confidence interval, N=number, P=p-value, BMI=body mass index

The validity of our dataset is also substantiated by genome-wide association studies (GWAS) that we conducted in EPIC4ND using logistic regression analyses of AD, PD, and ALS (additive model, adjusted for the first four principal components, sex, and age at baseline). These analyses confirmed that the vast majority of EPIC4ND participants is of European descent (Supplementary Fig. 1). These analyses revealed at least nominally significant effects of various well-established ND risk loci [61–64], including *APOE* rs429358 (OR=4.02, p=4.73E-24) for AD, *GBA* rs35749011 (OR 2.44, p=4.81–03), *SNCA* rs356182 (OR 0.80, p=6.43–03) and *MAPT* rs62053943 (OR=0.79, p=3.58E-02) for PD, and *C9ORF72* rs2453555 (OR=1.36,p=1.13E-02) for ALS (Supplementary Table 7, Supplementary Fig. 2). Several other known GWAS risk loci showed nominally significant effect estimates pointing into the same direction with 95% confidence intervals overlapping between the previous GWAS and the EPIC4ND GWAS. Exceptions were observed for *MS4A4E* rs1582763 (AD risk), for which effects estimate directions were inverse, and for *CLCN3* rs62333164 (ALS risk), for which the effect estimate was slightly stronger in the EPIC4ND study (Supplementary Table 7). Furthermore, polygenic risk scores (PGS) for AD, PD, and ALS generated in the EPIC4ND case-cohort based on the established independent genome-wide significant ($\alpha = 5 \times 10^{-8}$) GWAS risk SNPs [61–64] revealed strong associations with AD, PD, and ALS status. Specifically, we observed an OR of 2.04, 1.46, and 1.43 for AD, PD, and ALS, respectively, per one SD increase of the PGS (Supplementary Table 8). This supports the robustness of the case asertainment as well as the quality of the genetic data.

Furthermore, initial findings on plasma metabolites generated using mass spectrometry and on metals in erythrocytes have recently been published using overlapping, nested case–control studies for PD [16, 17, 19] and ALS [15].

Comparison with other resources, strengths and limitations

This population-based case-cohort study is unique in the ND field for its study design with the availability of pre-disease blood samples, an extensive follow-up period of up to 30 years, and the new, in-depth measurements of multi-layered molecular data. Furthermore, EPIC4ND ensures high diagnostic accuracy by validating dementia, AD and PD cases through a re-review of medical records by expert neurologists following an initial case ascertainment via data linkage. This is in contrast to most mega-cohort studies with baseline biomaterial available, such as the UK Biobank [65], that rely solely on data linkage to study incident neurodegenerative diseases. Moreover, all EPIC4ND participants underwent extensive phenotyping and risk factor assessment, in particular diet, at baseline, yielding a comprehensive range of lifestyle and medical variables. In addition and despite being a multi center study, both the phenotypic characterization and the processes for retrieving and storing blood samples have been highly standardized and harmonized, minimizing between-center heterogeneity.

When compared to other prospective cohorts with data on incident neurodegenerative disease cases and non-genetic omics data available, the UK Biobank [65] is the one of the best-established resources. It provides proteomic data for approximately 52,000 participants using the Olink Explore 3072 platform, currently covering around 3000 protein markers [66] (versus approximately 7000 protein

markers available in EPIC4ND). While the UKB substantially surpasses EPIC4ND in overall sample size, the case-cohort design applied to EPIC4ND resulted in comparable numbers of incident cases of AD (n in UKB ≈ 600), PD (n ≈ 800), and ALS (n ≈ 200) [29]. Importantly, the UKB (recruited 2006–2010) has a much shorter follow-up period than EPIC4ND (recruited 1991–2000). Other omics-based cohorts with blood-based proteomic and other omic data investigating neurodegenerative diseases include deCODE (total n = 36,000) [67, 68], ARIC (total cohort size: n ≈ 11,000) [38], and AGES (total n ≈ 5000) [39]. Although these cohorts often have larger overall sample sizes (for some studies this is also true for the number of dementia/AD cases), we note that EPIC4ND includes substantially more incident PD and ALS cases. As a matter of fact, ALS has not typically been investigated in these cohorts due to the overall very low case numbers. Additionally, EPIC4ND is distinguished by the availability of multi-omics data available for each participant of the core dataset.

Notwithstanding, our study has several weaknesses: First, although we aimed for the highest possible quality in ascertainment of AD, PD, and ALS cases within our study, the accuracy of the diagnosis ultimately lies in the responsibility of the diagnosing or treating physician, and there may be some variability across centers. Second, despite the above described multi-layered comprehensive linkage for case ascertainment, a limitation of this study is the likely lower sensitivity of health record-coded dementia/AD compared to PD and ALS [69]. This limitation, which may apply to many registry-based epidemiological studies, likely leads to misclassification of some AD cases in the subcohort as non-cases. Thus, for dementia/AD, observed effect size estimates likely represent conservative (under)estimates of the truly underlying associations. At the same time, while PD and ALS diagnoses are likely more accurate than that of AD, they are comparatively infrequent diseases, so any misclassification may have a proportionally larger consequence on the effect estimates than in common conditions like AD. Third, as in most epidemiological studies, some sources of bias may exist: A potential limitation of our study is selection bias, which can occur in prospective cohorts if participation or retention depends on both the exposure and the outcome [70]. In EPIC4ND this risk is likely mitigated given the random sampling of the subcohort where recruitment was irrespective of future case status, the prospective study design, and the diverse multi-layered case ascertainment methods. Furthermore, competing risk due to death from causes other than the outcome of interest is a common issue in survival analyses, especially in late-onset diseases such as ND. In EPIC4ND, we usually account for this by using cause-specific hazard models, which treat competing events (such as death from other causes than ND) as right-censored at the time of occurrence providing valid estimates of the association between exposures and ND incidence [71]. In addition, despite studying late-onset diseases we acknowledge the potential for survival bias (left-truncation) given that our baseline cohort consists of mostly middle-aged individuals who must haved survived until the age of study recruitment. Hence, severe ND cases with early onset may have been missed. By using age as the underlying time scale and implementing left-truncated Cox proportional hazards models, participants only enter the risk sets at their age of recruitment, thereby comparing individuals of the same age and mitigating the impact of delayed entry on the corresponding hazard ratios [72]. Notwithstanding, some risk of bias remains, but this is the case for nearly any epidemiological study. Fourth, while we aimed to obtain high-quality diagnostic information standardized evaluations of specific disease phenotypes (e.g., initial symptoms) and of disease progression (e.g., severity or therapy response) are largely missing. Lastly, EPIC is based in Europe and as demonstrated above the vast majority of participants are of European descent. As a result, although the ascertainment of a pan-European cohort is a strength in terms of generalizability to European populations, the generalizability of findings to other ethnic groups and populations is limited, highlighting the urgent need for future studies including more diverse populations.

Conclusions

EPIC4ND is a unique resource for identifying novel molecular biomarkers for a future onset of AD, PD, and ALS and for investigating biomarkers and pathways shared across ND and other diseases. Additionally, it enables the investigation of the interplay and interactions among multiple omics domains. Its validity is supported by associations with established lifestyle and genetic risk factors. Similar studies are urgently needed to address the critical need for identifying individuals at high risk of ND, facilitating the development of targeted preventive and early therapeutic interventions.

Data access

Data can be accessed by external researchers provided the proposed project is approved by the EPIC4ND working group and the EPIC steering committee. Potential collaborators can contact the EPIC4ND working group chair, Christina Lill (christina.lill@uni-muenster.de), EPIC administrator Sherry Morris (epicadmin@imperial.ac.uk), and EPIC PI Elio Riboli (e.riboli@imperial.ac.uk). Further

information regarding data access can be found at: https://epic.iarc.fr/access/. Data dictionaries for EPIC are available at: https://epic.iarc.fr/working-documents/.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10654-026-01430-1.

Acknowledgements The authors thank all participants of the project and the countless scientists who have contributed to the EPIC cohort over the past 30 years. We sincerely thank Bertrand Hemon for his assistance with data management, Tanja Wesse for technical support, and Prof. Paolo Vineis, Prof. Beate Ritz, Prof. Martin Wolkewitz, Prof. Klaus Berger and Prof. Valentina Gallo for helpful discussions.

Author contributions C.M.L., M.G., and E.R., conceived the study with help from S.G., L.M., L.B., and P.F.; C.M.L., J.H., K.S.-B., V.V., J.M.H., A.B., V.D., O.M., M.G., D.P., M.T., J.S., M.G., S.M. C.-Y., S.M., N. C.-C., S.E., S.S., A.J., T.Y.N.T., R.K., G.S., R.T., N.W., A.L.B., H.Z., A.F., A.E., L.B., R.V., L.M., G.M., C.S., S.P., V.K., P.F., M.G., and E.R. carried out data collection; C.M.L., J.H., O.O., K. S.-B., V.V., Y.Z., F.A., L.D., O.R., A.E., L.B., P.F., E.R. and M.G. carried out analysis and/or interpretation of data. C.M.L. and J.H. drafted the manuscript; all authors critically revised the manuscript for intellectual content. All authors read and approved the final manuscript of the paper.

Funding The updates of the case ascertainments, generation and processing of the molecular data were funded by the Michael J Fox Foundation (#008994 to C.M.L. and E.R.), the Cure Alzheimer's Fund (to C.M.L. and L.B.), the 'CReATe- Clinical Research in ALS and Related Disorders for Therapeutic Development' Consortium (to C.M.L. and L.B.), a grant from the EU Joint Programme – Neurodegenerative Disease Research (JPND2021-650–289, coordinator: C.M.L.), and the Deutsche Forschungsgemeinschaft (DFG; LI 2654/3–1, to C.M.L.). The Dutch partner of the JPND project was funded by ZonMw and Alzheimer Nederland. The case ascertainment for the dementia cases in Granada was funded by the Instituto de Investigación Biosanitaria ibs.GRANADA (INTRAIBS-2020–06). C.M. Lill was supported by the Heisenberg program of the DFG (DFG; LI 2654/4–1). See Supplementary Material for additional funding information and a disclaimer.

Declarations

Competing interests H.Z. has served at scientific advisory boards and/or as a consultant for Abbvie, Acumen, Alector, Alzinova, ALZpath, Amylyx, Annexon, Apellis, Artery Therapeutics, AZTherapies, Cognito Therapeutics, CogRx, Denali, Eisai, Enigma, LabCorp, Merry Life, Nervgen, Novo Nordisk, Optoceutics, Passage Bio, Pinteon Therapeutics, Prothena, Quanterix, Red Abbey Labs, reMYND, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures sponsored by Alzecure, BioArctic, Biogen, Cellectricon, Fujirebio, Lilly, Novo Nordisk, Roche, and WebMD, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. None of the other authors reports any relevant conflict of interests.

Ethical approval The study was approved by the ethical committee of the International Agency for Research on Cancer (IARC) and by the ethical review boards of participating study centers.

Consent to participate All participants provided written informed consent.

Consent to publish Not applicable.

References

1. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26(Suppl 1):S6–14. https://doi.org/10.1093/ije/26.suppl_1.s6.
2. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002;5(6B):1113–24. https://doi.org/10.1079/PHN2002394.
3. Slimani N, Kaaks R, Ferrari P, et al. European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. Public Health Nutr. 2002;5(6B):1125–45. https://doi.org/10.1079/PHN2002395.
4. Agudo A, Amiano P, Barcos A, et al. Dietary intake of vegetables and fruits among adults in five regions of Spain. Eur J Clin Nutr. 1999;53(3):174–80. https://doi.org/10.1038/sj.ejcn.1600694.
5. Bingham SA, Welch AA, McTaggart A, et al. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. Public Health Nutr. 2001;4(3):847–58. https://doi.org/10.1079/phn2000102.
6. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. Ann Nutr Metab. 1999;43(4):205–15. https://doi.org/10.1159/000012787.
7. Beulens JWJ, Monninkhof EM, Verschuren WMM, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170–8. https://doi.org/10.1093/ije/dyp217.
8. InterAct Consortium, Langenberg C, Sharp S, et al. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia*. 2011;54(9):2272–2282. https://doi.org/10.1007/s00125-011-2182-9
9. Danesh J, Saracci R, Berglund G, et al. EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. Eur J Epidemiol. 2007;22(2):129–41. https://doi.org/10.1007/s10654-006-9096-8.
10. Lopes EW, Chan SSM, Song M, et al. Lifestyle factors for the prevention of inflammatory bowel disease. Gut. 2022. https://doi.org/10.1136/gutjnl-2022-328174.
11. Gallo V, Bueno-De-Mesquita HB, Vermeulen R, et al. Smoking and risk for amyotrophic lateral sclerosis: analysis of the EPIC cohort. Ann Neurol. 2009;65(4):378–85. https://doi.org/10.1002/ana.21653.
12. Gallo V, Wark PA, Jenab M, et al. Prediagnostic body fat and risk of death from amyotrophic lateral sclerosis: the EPIC cohort. Neurology. 2013;80(9):829–38. https://doi.org/10.1212/WNL.0b013e3182840689.

13. Gallo V, Brayne C, Forsgren L, et al. Parkinson's Disease Case Ascertainment in the EPIC cohort: The NeuroEPIC4PD study. Neurodegener Dis. 2015;15(6):331–8. https://doi.org/10.1159/000381857.
14. Gallo V, Vanacore N, Bueno-de-Mesquita HB, et al. Physical activity and risk of Amyotrophic Lateral Sclerosis in a prospective cohort study. Eur J Epidemiol. 2016;31(3):255–66. https://doi.org/10.1007/s10654-016-0119-9.
15. Peters S, Broberg K, Gallo V, et al. Blood metal levels and amyotrophic lateral sclerosis risk: a prospective cohort. Ann Neurol. 2021;89(1):125–33. https://doi.org/10.1002/ana.25932.
16. Zhao Y, Ray A, Broberg K, et al. Prediagnostic blood metal levels and the risk of Parkinson's disease: a large european prospective cohort. Mov Disord. 2023;38(12):2302–7. https://doi.org/10.1002/mds.29602.
17. Zhao Y, Lai Y, Konijnenberg H, et al. Association of coffee consumption and prediagnostic caffeine metabolites with incident parkinson disease in a population-based cohort. Neurology. 2024;102(8):e209201. https://doi.org/10.1212/WNL.0000000000209201.
18. Gröninger M, Sabin J, Kaaks R, et al. Associations of milk, dairy products, calcium and vitamin D intake with risk of developing Parkinson´s disease within the EPIC4ND cohort. Eur J Epidemiol. 2024;39(11):1251–65. https://doi.org/10.1007/s10654-024-01183-9.
19. Zhao Y, Lai Y, Darweesh SKL, et al. Gut microbial metabolites and future risk of Parkinson's disease: a metabolome-wide association study. Mov Disord. 2025;40(3):556–60. https://doi.org/10.1002/mds.30054.
20. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. The Lancet Public Health. 2022;7(2):e105–25. https://doi.org/10.1016/S2468-2667(21)00249-8.
21. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18(5):459–80. https://doi.org/10.1016/S1474-4422(18)30499-X.
22. Arthur KC, Calvo A, Price TR, Geiger JT, Chiò A, Traynor BJ. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat Commun. 2016;7:12408. https://doi.org/10.1038/ncomms12408.
23. Rafii MS. Preclinical Alzheimer's disease therapeutics. J Alzheimers Dis. 2014;42(4):S545–9. https://doi.org/10.3233/JAD-141482.
24. Palmqvist S, Tideman P, Cullen N, et al. Prediction of future Alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. Nat Med. 2021;27(6):1034–42. https://doi.org/10.1038/s41591-021-01348-z.
25. Cullen NC, Leuzy A, Janelidze S, et al. Plasma biomarkers of Alzheimer's disease improve prediction of cognitive decline in cognitively unimpaired elderly populations. Nat Commun. 2021;12(1):3555. https://doi.org/10.1038/s41467-021-23746-0.
26. Salvadó G, Janelidze S, Bali D, et al. Plasma Phosphorylated Tau 217 to identify Preclinical Alzheimer disease. JAMA Neurol. 2025;82(11):1122–34. https://doi.org/10.1001/jamaneurol.2025.3217.
27. Cai H, Pang Y, Fu X, Ren Z, Jia L. Plasma biomarkers predict Alzheimer's disease before clinical onset in Chinese cohorts. Nat Commun. 2023;14(1):6747. https://doi.org/10.1038/s41467-023-42596-6.
28. Valletta M, Vetrano DL, Gregorio C, et al. Blood biomarkers of Alzheimer's disease and progression across different stages of cognitive decline in the community. Nat Commun. 2025;16(1):10412. https://doi.org/10.1038/s41467-025-66728-2.
29. Buonocore J, Fratto E, Arcuri F, et al. Plasma NfL and GFAP in the preclinical stages of neurodegenerative diseases: insights from the UK Biobank. J Neurol. 2025;272(12):755. https://doi.org/10.1007/s00415-025-13498-y.
30. Grande G, Valletta M, Rizzuto D, et al. Blood-based biomarkers of Alzheimer's disease and incident dementia in the community. Nat Med. 2025;31(6):2027–35. https://doi.org/10.1038/s41591-025-03605-x.
31. Shadyab AH, Zhang B, LaCroix AZ, et al. Plasma p-tau217 and incident mild cognitive impairment and dementia in older women: 25-year prospective study in The Women's Health Initiative Memory Study. *medRxiv*. Published online November 11, 2025. https://doi.org/10.1101/2025.10.30.25339146
32. Okuzumi A, Hatano T, Matsumoto G, et al. Propagative α-synuclein seeds as serum biomarkers for synucleinopathies. Nat Med. 2023;29(6):1448–55. https://doi.org/10.1038/s41591-023-02358-9.
33. Yan S, Jiang C, Janzen A, et al. Neuronally derived extracellular vesicle α-Synuclein as a serum biomarker for individuals at risk of developing Parkinson Disease. JAMA Neurol. 2024;81(1):59–68. https://doi.org/10.1001/jamaneurol.2023.4398.
34. Hatano T, Okuzumi A, Matsumoto G, Tsunemi T, Hattori N. α-Synuclein: a promising biomarker for Parkinson's disease and related disorders. J Mov Disord. 2024;17(2):127–37. https://doi.org/10.14802/jmd.24075.
35. Obara K, Ito D, Nilsson C, et al. Diagnostic and prognostic value of blood and cerebrospinal fluid biomarkers in amyotrophic lateral sclerosis: a systematic review and meta-analysis. Eur J Neurol. 2025;32(10):e70382. https://doi.org/10.1111/ene.70382.
36. Bjornevik K, O'Reilly EJ, Molsberry S, et al. Prediagnostic neurofilament light chain levels in amyotrophic lateral sclerosis. Neurology. 2021;97(15):e1466–74. https://doi.org/10.1212/WNL.0000000000012632.
37. Walker KA, Chen J, Zhang J, et al. Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. Nat Aging. 2021;1(5):473–89. https://doi.org/10.1038/s43587-021-00064-0.
38. Walker KA, Chen J, Shi L, et al. Proteomics analysis of plasma from middle-aged adults identifies protein markers of dementia risk in later life. *Sci Transl Med*. 2023;15(705). https://doi.org/10.1126/scitranslmed.adf5681
39. Frick EA, Emilsson V, Jonmundsson T, et al. Serum proteomics reveal APOE-ε4-dependent and APOE-ε4-independent protein signatures in Alzheimer's disease. Nat Aging. 2024;4(10):1446–64. https://doi.org/10.1038/s43587-024-00693-1.
40. Guo Y, You J, Zhang Y, et al. Plasma proteomic profiles predict future dementia in healthy adults. Nat Aging. 2024;4(2):247–60. https://doi.org/10.1038/s43587-023-00565-0.
41. Gan YH, Ma LZ, Zhang Y, et al. Large-scale proteomic analyses of incident Parkinson's disease reveal new pathophysiological insights and potential biomarkers. Nat Aging. 2025;5(4):642–57. https://doi.org/10.1038/s43587-025-00818-0.
42. O'Brien KM, Lawrence KG, Keil AP. The case for case-cohort: an applied epidemiologist's guide to reframing case-cohort studies to improve usability and flexibility. Epidemiology. 2022;33(3):354–61. https://doi.org/10.1097/EDE.0000000000001469.
43. Andreu-Reinón ME, Gavrila D, Chirlaque MD, et al. Ascertainment of dementia cases in the Spanish European Prospective Investigation into Cancer and Nutrition-Murcia cohort. Neuroepidemiology. 2019;52(1–2):63–73. https://doi.org/10.1159/000493209.
44. Andreu-Reinón ME, Huerta JM, Gavrila D, et al. Incidence of dementia and associated factors in the EPIC-Spain dementia cohort. J Alzheimers Dis. 2020;78(2):543–55. https://doi.org/10.3233/JAD-200774.

45. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry. 1988;51(6):745–52. https://doi.org/10.1136/jnnp.51.6.745.
46. Harding Z, Wilkinson T, Stevenson A, et al. Identifying Parkinson's disease and parkinsonism cases using routinely collected healthcare data: A systematic review. PLoS ONE. 2019;14(1):e0198736. https://doi.org/10.1371/journal.pone.0198736.
47. Marin B, Logroscino G, Boumédiene F, et al. Clinical and demographic factors and outcome of amyotrophic lateral sclerosis in relation to population ancestral origin. Eur J Epidemiol. 2016;31(3):229–45. https://doi.org/10.1007/s10654-015-0090-x.
48. Nguyen TTH, Fournier A, Courtois É, et al. Use of β-adrenoreceptor drugs and Parkinson's disease incidence in women from the French E3N cohort study. J Parkinsons Dis. 2025;15(4):789–804. https://doi.org/10.1177/1877718X251330993.
49. Clavel-Chapelon F, van Liere MJ, Giubout C, et al. E3N, a French cohort study on cancer risk factors. Eur J Cancer Prev. 1997;6(5):473–8. https://doi.org/10.1097/00008469-199710000-00007.
50. Lövheim H, Olsson J, Weidung B, et al. Interaction between cytomegalovirus and herpes simplex virus type 1 associated with the risk of Alzheimer's Disease development. J Alzheimers Dis. 2018;61(3):939–45. https://doi.org/10.3233/JAD-161305.
51. Hong S, Prokopenko D, Dobricic V, et al. Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD Multimodal Biomarker Discovery dataset. Transl Psychiatry. 2020;10(1):403. https://doi.org/10.1038/s41398-020-01074-z.
52. Dobricic V, Schilling M, Schulz J, et al. Differential microRNA expression analyses across two brain regions in Alzheimer's disease. Transl Psychiatry. 2022;12(1):352. https://doi.org/10.1038/s41398-022-02108-4.
53. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika. 1986;73(1):1–11. https://doi.org/10.1093/biomet/73.1.1.
54. Barlow WE. Robust variance estimation for the case-cohort design. Biometrics. 1994;50(4):1064–72.
55. Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. J Clin Epidemiol. 1999;52(12):1165–72. https://doi.org/10.1016/s0895-4356(99)00102-x.
56. Onland-Moret NC, van der A DL, van der Schouw YT, et al. Analysis of case-cohort data: a comparison of different methods. *J Clin Epidemiol*. 2007;60(4):350–355. https://doi.org/10.1016/j.jclinepi.2006.06.022
57. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. Stat Biosci. 2009;1(1):32. https://doi.org/10.1007/s12561-009-9001-6.
58. Gallo V, Vineis P, Cancellieri M, et al. Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol. 2019;48(3):912–25. https://doi.org/10.1093/ije/dyy230.
59. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020;396(10248):413–46. https://doi.org/10.1016/S0140-6736(20)30367-6.
60. Noyce AJ, Bestwick JP, Silveira-Moriyama L, et al. Meta-analysis of early nonmotor features and risk factors for Parkinson disease. Ann Neurol. 2012;72(6):893–901. https://doi.org/10.1002/ana.23687.
61. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019;18(12):1091–102. https://doi.org/10.1016/S1474-4422(19)30320-5.
62. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12):1452–8. https://doi.org/10.1038/ng.2802.
63. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet. 2022;54(4):412–36. https://doi.org/10.1038/s41588-022-01024-z.
64. van Rheenen W, van der Spek RAA, Bakker MK, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat Genet. 2021;53(12):1636–48. https://doi.org/10.1038/s41588-021-00973-1.
65. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9. https://doi.org/10.1038/s41586-018-0579-z.
66. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622(7982):329–38. https://doi.org/10.1038/s41586-023-06592-6.
67. Eldjarn GH, Ferkingstad E, Lund SH, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. Nature. 2023;622(7982):348–58. https://doi.org/10.1038/s41586-023-06563-x.
68. Eldjarn GH, Ferkingstad E, Lund SH, et al. Correction to: Large-scale plasma proteomics comparisons through genetics and disease associations (Nature, (2023), 622, 7982, (348–358), https://doi.org/10.1038/s41586-023-06563-x). Nature.Nature Research. 2024;630(8015):E3. 10.1038/s41586–024–07549-z
69. Lusk JB, Choi S, Clark AG, et al. Dementia and Parkinson's disease diagnoses in electronic health records vs Medicare claims data: a study of 101,980 linked patients. BMC Neurol. 2023;23(1):325. https://doi.org/10.1186/s12883-023-03361-w.
70. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615–25. https://doi.org/10.1097/01.ede.0000135174.63482.43.
71. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009;170(2):244–56. https://doi.org/10.1093/aje/kwp107.
72. Thiébaut ACM, Bénichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. Stat Med. 2004;23(24):3803–20. https://doi.org/10.1002/sim.2098.

Authors and Affiliations

Christina M. Lill[1,2] · Jan Homann[1] · Olena Ohlei[1] · Karl Smith-Byrne[3] · Vivian Viallon[4] · Jose Maria Huerta[5,6] · Fanny Artaud[7] · Yujia Zhao[8] · Abigail Britten[9] · Laura Deecke[1] · Valerija Dobricic[10] · Olatz Mokoroa[11,12] · Marcela Guevara[13,14,15] · Dafina Petrova[16,17,14] · Miles Trupp[18] · Jara Sabin[19,20] · Mareike Gröninger[19] · Sandra M. Colorado-Yohar[5,6,21] · Sonia Alonso-Martin[22] · Natalia Cabrera-Castro[5,6] · Sergiu Groppa[23] · Oliver Robinson[2,24] · Johnni Hansen[25] · Sture Eriksson[26] · Sabina Sieri[27] · Ana Jimenez[28,29] · Tammy Y. N. Tong[3] · Rudolph Kaaks[19] · Gianluca Severi[7,30] · Ruth Travis[3] · Nick Wareham[9] · Andréa L. Benedet[31] · Henrik Zetterberg[31,32,33,34,35,36] · Andre Franke[37] · Alexis Elbaz[7] · Lars Bertram[10] · Roel Vermeulen[8,38] · Lefkos Middleton[2] · Giovanna Masala[39] · Carlotta Sacerdote[40] · Susan Peters[8] · Verena Katzke[19] · Pietro Ferrari[4] · Marc J. Gunter[24] · Elio Riboli[24]

✉ Christina M. Lill
christina.lill@uni-muenster.de; clill@ic.ac.uk

✉ Elio Riboli
e.riboli@imperial.ac.uk

1 Translational Epidemiology Unit, Institute of Epidemiology and Social Medicine, University of Münster, Domagkstr. 3, Münster, Germany

2 Ageing and Epidemiology Unit (AGE), School of Public Health, Imperial College London, London, UK

3 Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

4 International Agency for Research on Cancer (IARC/WHO), Nutrition and Metabolism Branch, Lyon, France

5 Department of Epidemiology, Murcia Regional Health Council-IMIB, Murcia, Spain

6 CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

7 Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Villejuif, France

8 Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

9 Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK

10 Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Lübeck, Germany

11 Epidemiology and Public Health Area, Biodonostia Health Research Institute, San Sebastián, Spain

12 Sub-Directorate for Public Health and Addictions of Gipuzkoa, Ministry of Health of the Basque Government, San Sebastián, Spain

13 Instituto de Salud Pública y Laboral de Navarra, Pamplona, Spain

14 Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

15 Navarra Institute for Health Research (IdiSNA), Pamplona, Spain

16 Escuela Andaluza de Salud Pública (EASP), Granada, Spain

17 Instituto de Investigación Biosanitaria Ibs.GRANADA, Granada, Spain

18 Department of Clinical Sciences, Neurosciences, Umeå University, Umeå, Sweden

19 Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

20 1st Medical Clinic, University Hospital Mannheim, Mannheim, Germany

21 9 Enero Research Group On Demography and Health, National Faculty of Public Health, University of Antioquia, Medellin, Colombia

22 Stem Cells and Aging Group Bioengineering Area, BioGipuzkoa (BioDonostia) Health Research Institute, San Sebastián, Spain

23 Department of Neurology, Saarland University Clinic, Saarland, Germany

24 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

25 Danish Cancer Institute, Danish Cancer Society, Copenhagen, Denmark

26 Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

27 Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy

28 Sub Directorate for Public Health and Addictions of Gipuzkoa, Ministry of Health of the Basque Government, San Sebastian, Spain

29 Epidemiology of Chronic and Communicable Diseases Group, BioGipuzkoa (BioDonostia) Health Research Institute, San Sebastián, Spain

30 Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence, Florence, Italy

31 Department of Psychiatry and Neurochemistry, University of Gothenburg, Molndal, Sweden

32 Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

33 Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, UK

34 UK Dementia Research Institute at UCL, London, UK

35 Hong Kong Center for Neurodegenerative Diseases, Clear Water Bay, Hong Kong, China

[36] Wisconsin Alzheimer's Disease Research Center, School of Medicine and Public Health, University of Wisconsin, University of Wisconsin-Madison, Madison, WI, USA

[37] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany

[38] University Medical Centre Utrecht, Utrecht, The Netherlands

[39] Clinical Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Florence, Italy

[40] Department of Health Sciences, University of Eastern Piedmont, Novara, Italy