Identifying health conditions associated with Alzheimer's disease up to 15 years before diagnosis: an agnostic study of French and British health records

Thomas Nedelec, Baptiste Couvy-Duchesne, Fleur Monnet, Timothy Daly, Manon Ansart, Laurène Gantzer, Béranger Lekens, Stéphane Epelbaum, Carole Dufouil, Stanley Durrleman**

Summary

Background The identification of modifiable risk factors for Alzheimer's disease is paramount for early prevention and the targeting of new interventions. We aimed to assess the associations between health conditions diagnosed in primary care and the risk of incident Alzheimer's disease over time, up to 15 years before a first Alzheimer's disease diagnosis.

Methods In this agnostic study of French and British health records, data from 20 214 patients with Alzheimer's disease in the UK and 19 458 patients with Alzheimer's disease in France were extracted from The Health Improvement Network database. We considered data recorded from Jan 1, 1996, to March 31, 2020 in the UK and from Jan 4, 1998, to Feb 20, 2019, in France. For each Alzheimer's disease case, a control was randomly assigned after matching for sex and age at last visit. We agnostically tested the associations between 123 different diagnoses of the International Classification of Diseases, 10th revision, extracted from health records, and Alzheimer's disease, by running a conditional logistic regression to account for matching of cases and controls. We focused on three time periods before diagnosis of Alzheimer's disease, to separate risk factors from early symptoms and comorbidities.

Findings Unadjusted odds ratios (ORs) and 95% CIs for the association between Alzheimer's disease and various health conditions were estimated, and p values were corrected for multiple comparisons. In both the British and French studies, ten health conditions were significantly positively associated with increased Alzheimer's disease risk, in a window of exposure from 2–10 years before Alzheimer's disease diagnosis, comprising major depressive disorder (UK OR 1·34, 95% CI 1·23–1·46; France OR 1·73, 1·57–1·91), anxiety (UK OR 1·36, 1·25–1·47; France OR 1·50, 1·36–1·65), reaction to severe stress and adjustment disorders (UK OR 1·40, 1·24–1·59; France OR 1·83, 1·55–2·15), hearing loss (UK OR 1·19, 1·11–1·28; France OR 1·51, 1·21–1·89), constipation (UK OR 1·31, 1·22–1·41; France OR 1·59, 1·44–1·75), spondylosis (UK OR 1·26, 1·14–1·39; France OR 1·62, 1·44–1·81), abnormal weight loss (UK OR 1·47, 1·33–1·63; France OR 1·88, 1·56–2·26), malaise and fatigue (UK OR 1·23, 1·14–1·32; France OR 1·59, 1·46–1·73), memory loss (UK OR 7·63, 6·65–8·76; France OR 4·41, 3·07–6·34), and syncope and collapse (UK OR 1·23, 1·10–1·37; France OR 1·57, 1·26–1·96). Depression was the first comorbid condition associated with Alzheimer's disease, appearing at least 9 years before the first clinical diagnosis, followed by anxiety, constipation, and abnormal weight loss.

Interpretation These results from two independent primary care databases provide new evidence on the temporality of risk factors and early signs of Alzheimer's disease that are observable at the general practitioner level. These results could guide the implementation of new primary and secondary prevention policies.

Funding Agence Nationale de la Recherche.

Lancet Digit Health 2022; 4: e169–78

See **Comment** page e150

*Contributed equally

Paris Brain Institute, ICM, INRIA, INSERM, CNRS, Sorbonne Université, Paris, France (T Nedelec PhD, B Couvy-Duchesne PhD, M Ansart PhD, S Epelbaum PhD, S Durrleman PhD); **Department of Neurology, Assistance Publique-Hôpitaux de Paris, Paris, France** (S Epelbaum); **Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD, Australia** (B Couvy-Duchesne); **CEGEDIM R&D, Boulogne-Billancourt, France** (F Monnet MSc, L Gantzer MSc, B Lekens MSc); **Department of Philosophy, Sorbonne Université, Paris, France** (T Daly MPhil); **UMR 1219, Bordeaux Population Health, INSERM, CIC1401-EC, Bordeaux Université, Bordeaux, France** (C Dufouil PhD); **F-33000, Pole de Sante Publique Centre Hospitalier Universitaire de Bordeaux, Bordeaux, France** (C Dufouil)

Correspondence to:
Dr Carole Dufouil, UMR 1219, Bordeaux Population Health, Inserm, CIC1401-EC, Bordeaux Université, Bordeaux 33076, France
carole.dufouil@inserm.fr

Introduction

Alzheimer's disease is thought to account for 60–70% of dementia cases worldwide, and is thus one of the principal health challenges of the 21st century.[1] Dementia affected 10·5 million Europeans in 2015,[2] and projections suggest that 13·5 million Europeans will be affected in 2030.[2] It is estimated that 4·5% of people older than 65 years currently have Alzheimer's disease in Europe (and 6·4% have dementia).[2] Alzheimer's disease has major consequences at the individual and societal levels because it requires a high degree of social care.[3]

The neurodegenerative diseases responsible for dementia (and Alzheimer's disease in particular) are progressive and develop over decades before becoming disabling.[4,5] Attempts to develop an effective disease-modifying treatment for patients with symptomatic Alzheimer's disease have been unsuccessful. There has, therefore, been a shift towards early interventions, to maximise the therapeutic window through primary prevention measures acting on actionable risk factors, or secondary prevention measures to slow disease progression through early therapeutic interventions. The

Research in context

Evidence before this study

We searched PubMed from inception up to May 20, 2021, for reports in English with the terms "Alzheimer's disease", "case-control", and "prodromal", "prediagnostic", or "pre-diagnostic", yielding 194 reports. We screened these case-control or cohort studies with a history of at least 5 years before diagnosis of Alzheimer's disease for content and did not find any papers on the prodromal stage of Alzheimer's disease that agnostically addressed the most frequent health conditions observable by general practitioners.

Added value of this study

This study provides a modern extensive data-driven survey of potential risk factors and early symptomatic expression for Alzheimer's disease in primary care, observable by general practitioners, and constitutes a first attempt to establish when these conditions appear during the course of the disease up to 15 years before the first diagnosis. After controlling for multiple testing, ten of 123 tested health conditions were significantly associated with a higher risk of developing Alzheimer's disease 2–10 years before diagnosis. Only anxiety, depression, constipation, spondylosis, and memory loss were associated at shown by multivariate analysis.

Implications of all the available evidence

This work is important to better inform general practitioners and patients about the risk factors of Alzheimer's disease, one of the diseases in which primary care is the most difficult to manage. It is equally important for policy makers and drug developers to better design primary and secondary prevention measures.

key to the implementation of such preventive measures is a data-driven understanding of the complex dynamics of the presymptomatic period at the point of care.

The most recent *Lancet* Commission on dementia prevention, intervention, and care[6] added three new modifiable risk factors (excessive alcohol consumption, head injury, and air pollution) to the initial list of nine risk factors[3] (low level of education, hypertension, hearing impairment, smoking, obesity, depression, physical inactivity, diabetes, and infrequent social contact). The report concluded that modifying these 12 risk factors might prevent or delay dementia in up to 40% of cases (7% for low level of education and 8% for hearing loss). For Alzheimer's disease, another large meta-analysis of 243 observational prospective studies and 153 randomised controlled trials[7] identified ten modifiable risk factors, comprising diabetes, hyperhomocysteinaemia, poor BMI management, low level of education, hypertension in midlife, orthostatic hypotension, head trauma, poor cognitive activity, stress, and depression. All the studies stressed the importance of a specific feature of neurodegenerative diseases, which include a prodromal phase lasting up to 20 years before clinical diagnosis.[4,5] This prodromal phase makes it more difficult to separate possible early symptoms from causal risk factors when describing the natural course of the disease. It also highlights the urgent need for observational prospective studies with long follow-up periods[8–10] to investigate the long-term relationship between medical conditions and dementia onset.[4]

Epidemiological studies on the causes of Alzheimer's disease have focused on hypothesis-driven searches for potential risk factors in cohort studies on older people (mostly older than 60 years). By contrast, agnostic studies of risk factors test all possible associations exhaustively in a similar fashion to genome-wide association studies for finding genotype–phenotype associations. They require large samples for the detection of statistically significant signals for associations. The constitution of such large samples is made possible by access to large electronic health-record databases,[11] facilitating complex longitudinal analyses and the testing of multiple hypotheses in studies with sufficient power.

We set up and analysed two large nested case-control studies drawn from two large independent primary care databases in the UK and France, extracted from The Health Improvement Network (THIN) database. Through this extensive study, we sought to confirm some well known dementia risk factors and exhibit new candidates for consideration in future anti-dementia strategies.

Methods

Study design and participants

We used The Health Improvement Network (THIN) database,[12] a large standardised European database of fully anonymised and non-extrapolated electronic medical records collected from physicians by the company GERS SAS[12] and coded according to the International Classification of Diseases, 10th revision (ICD10) codes. The THIN database complies with all current European data protection laws (General Data Protection Regulation) and adheres to the Observational Medical Outcomes Partnership model.[12] The French data were collected from a pool of 2000 general practitioners,[13] and were representative of the French population in terms of age, sex, and living area. The UK data were collected from 400 general practices, representing around 6% of the UK population, selected for the THIN quality data-recording scheme with Vision practice management software.[14] Several reports have already shown that the electronically coded diagnoses in this database are representative of the UK general practice population in

terms of demographics and type of consultation.[12,15] For each patient, the diagnosis corresponding to each visit, the prescriptions made by the general practitioner, and all other diagnoses associated with these prescriptions were available. Data for educational level and social status were not provided, to preserve anonymity.

For each patient we had access to all the disease diagnoses made by the general practitioner, whether it was the main reason for the visit or the justification of a prescription. In the French cohort, we might not have had access to the full history of some patients if during the study these patients consulted other general practitioners who were not part of the considered panel. We defined past exposure to the health conditions considered according to ICD10 codes, which were provided directly in the French database. For the UK database, ICD10 codes were obtained by converting read codes according to the correspondence provided by the UK Clinical Terminology Centre[16] (appendix p 22). We used the first three characters of the code, defining the category of the disease. In this exploratory approach, we assessed the association between Alzheimer's disease and each of the health conditions defined by ICD10 codes recorded in more than 0·1% of visits per 1000 person-years in both countries (appendix p 1).

We considered data recorded from Jan 1, 1996, to March 31, 2020 in the UK and from Jan 4, 1998, to Feb 20, 2019 in France. We extracted two samples of patients diagnosed with Alzheimer's disease from the French and UK electronic health records in the THIN database (dementia cases according to ICD10 codes F00 and G30). In France, we extracted all patients with Alzheimer's disease and at least 2 years of follow-up (n=20 545). In the UK, we extracted an equivalent number of cases among patients with Alzheimer's disease with at least 2 years of follow-up (n=20 654). We further selected patients with at least 1 year of follow-up before diagnosis (n=19 458 in France and n=20 214 in the UK). Age at Alzheimer's disease onset was defined as age at the first record coded with an Alzheimer's disease diagnosis. For each country, a sample of control individuals was drawn, with no history of diagnosis of the neurodegenerative diseases coded as follows: Alzheimer's disease (F00 and G30); Parkinson's disease (G20); frontotemporal dementia (G31.0); dementia with Lewy bodies (G31.8); Huntington's disease (G10); or multiple sclerosis (G35). The control individuals were then matched with patients with Alzheimer's disease for age (SD 1 year) at last record in the database and sex. Patients with less than 2 years of follow-up were excluded (appendix p 5). A summary of the original data is provided (table 1).

Statistical analysis

We first used conditional logistic regression to analyse the association between each health condition corresponding to a specific ICD10 code and the diagnosis of Alzheimer's disease. On the basis of these initial cohorts, we ran several conditional logistic regressions. For the investigation of risk factors and symptoms, we considered several time frames preceding the diagnosis of Alzheimer's disease. We calculated ORs for three time periods between exposure to a particular health condition and dementia onset: 0–2 years, 2–10 years, and 10–15 years. Individuals acting as controls and patients with Alzheimer's disease were matched for age at their last visit and did not have to survive beyond the index date to be included. The index date for each case-control pair was defined as the first date of Alzheimer's disease diagnosis of the corresponding patient with Alzheimer's disease. For each analysis, we individually matched each patient with a control for sex and age at last visit (SD 1 year), and ensured that cases and controls had been observed over the same lifespan (appendix p 7). Ethical approval was obtained before carrying out this study (and for the use of datasets). The study was a retrospective analysis of secondary anonymised patient data only. For the UK database, data are only available to researchers carrying out approved medical research. Ethical approval was granted by the NHS South-East Multicentre Research Ethics Committee in 2003 (reference 03/01/073) for the establishment of the THIN database, and was updated in 2011. A further update was carried out and approval was granted in 2020 by the NHS South Central, Oxford C Research Ethics Committee (reference 20/SC/0011). For the French database, several audits were done by the Commission Nationale de l'Informatique et des Libertés, the French authority responsible for the protection of personal data. Data are available from GERS SAS for researchers who meet the criteria for access to confidential data.

See Online for appendix

P values were corrected for multiple comparisons using the Bonferroni method. Values of less than 0·0004 in two-tailed tests were considered significant (p=0·05 with Bonferroni correction for 123 potential exposures). We also report CIs adjusted for the corresponding confidence level (1 minus 0·0004). We estimated effect size independently from time to diagnosis and by time between health condition exposure and Alzheimer's disease diagnosis as follows: (0–2] years, (2–10] years, and (10–15] years.

We determined the independent contributions of the different health conditions to the risk of Alzheimer's disease by doing a conditional logistic regression including all the health conditions significantly and independently associated with an increase in Alzheimer's disease risk. Assuming causality, we calculated a combined population-attributable fraction associated with health conditions displaying significant association with Alzheimer's disease.

We tried to identify the window of exposure of greatest importance for the various health conditions by calculating the incidence of the health conditions concerned in people with Alzheimer's disease and their matched controls, and followed its change over time until the index

	Full analytical cohorts		People with ≥2 years of retrospective data		People with ≥10 years of retrospective data		People with ≥15 years of retrospective data	
	People with Alzheimer's disease (n=39 672)	Controls (n=39 672)	People with Alzheimer's disease (n=38 996)	Controls (n=38 996)	People with Alzheimer's disease (n=18 234)	Controls (n=18 234)	People with Alzheimer's disease (n=9180)	Controls (n=9180)
UK								
Total	20 214	20 214	19 940	19 940	13 064	13 064	7731	7731
Sex								
Female	14 107 (69%)	14 107 (69%)	13 924 (69%)	13 924 (69%)	9217 (70%)	9217 (70%)	5567 (72%)	5567 (72%)
Male	6107 (31%)	6107 (31%)	6016 (31%)	6016 (31%)	3847 (30%)	3847 (30%)	2164 (28%)	2164 (28%)
Age at index date, years	81 (76–86)	81 (76–86)	81 (76–86)	81 (76–86)	82 (76–86)	82 (76–86)	82 (77–86)	82 (77–86)
Median birth year	1926 (1921–1933)	1925 (1918–1933)	1926 (1921–1933)	1926 (1918–1933)	1928 (1922–1934)	1928 (1922–1935)	1929 (1924–1935)	1930 (1924–1936)
Person-years of data available before index date	13 (9–18)	10 (5–17)	13 (9–18)	11 (6–17)	16 (13–20)	16 (12–20)	19 (17–22)	19 (17–22)
Number of visits per year	7 (4–11)	7 (3–11)	7 (4–11)	7 (3–11)	7 (4–11)	7 (4–11)	7 (4–11)	7 (4–11)
France								
Total	19 458	19 458	19 056	19 056	5170	5170	1449	1449
Sex								
Women	12 442 (63%)	12 442 (63%)	12 186 (63%)	12 186 (63%)	3291 (63%)	3291 (63%)	906 (62%)	906 (62%)
Men	7016 (37%)	7016 (37%)	6870 (37%)	6870 (37%)	1879 (37%)	1879 (37%)	543 (38%)	543 (38%)
Age at index date, years	80 (75–85)	80 (75–85)	80 (75–85)	80 (75–85)	82 (77–86)	82 (77–86)	83 (77–86)	83 (77–86)
Birth year	1930 (1925–1935)	1931 (1925–1937)	1930 (1925–1935)	1931 (1926–1937)	1933 (1928–1938)	1934 (1929–1939)	1933 (1929–1939)	1934 (1930–1939)
Person-years of data available before index date	7 (4–10)	5 (3–8)	7 (4–10)	5 (4–8)	13 (11–15)	12 (11–14)	17 (16–18)	16 (16–17)
Number of visits per year	5 (3–8)	4 (2–7)	5 (3–8)	4 (2–7)	5 (3–8)	5 (2–7)	5 (3–7)	5 (3–7)

The data shown are number (%) or median (IQR).

Table 1: **Characteristics of patients with Alzheimer's disease and controls**

date, for all conditions found to be significantly associated with Alzheimer's disease in the time window (2–10 years). We merged the French-nested and the UK-nested case-control samples for this analysis. We assumed that incidence followed a Poisson distribution and calculated the associated 95% CIs. We added five additional comorbid conditions (sleep disorders, hypotension, disorders of urinary system, hypothyroidism, and fracture) associated with Alzheimer's disease when diagnosed in the 0–2 years preceding Alzheimer's disease diagnosis, which have also been reported as possible risk factors for dementia in previous studies.[6,17–19] We computed the change in incidence rate for each comorbid condition with time to Alzheimer's disease diagnosis in the Alzheimer's disease cohort. For each comorbid condition, we considered the first timepoint at which the 95% CI of the corresponding incidence rates between cases and controls were separate. All analyses were done with the StatsModel version 0.11.1 and ScikitLearn version 0.24.1 libraries in Python version 3.6.

Role of the funding source

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

Results

Our initial full analytical cohorts included 20 214 patients with a diagnosis of Alzheimer's disease in the UK and 19 458 patients with Alzheimer's disease in France (table 1).

The age and sex distribution of patients with Alzheimer's disease and their matched controls in the two countries are presented (table 1). Median age at Alzheimer's disease diagnosis was 81 years (76–86) years in the UK and 80 years (75–85) in France. There were more women than men in both countries. More follow-up data were available for the UK sample on average, with 13 person-years of data available before the index date in the full analytical cohort (versus 7 person-years on average in the French sample). The median number of visits per year was also higher in the UK than in France. 2 years before the index date, 19 940 patients with Alzheimer's disease were included in the UK study and 19 056 patients were included in the French study; 10 years before the index date, 13 064 patients were included in the UK study and 5170 patients were included in the French study; and 15 years before the index date, 7731 patients were included in the UK study and 1449 in the French study. Age at diagnosis was not dependent on the number of years of data available before the index

	Diagnosis within 0–2 years				Diagnosis within 2–10 years				Diagnosis within 10–15 years			
	UK		France		UK		France		UK		France	
	OR (corrected 95% CI)	Corrected p value	OR (corrected 95% CI)	Corrected p value	OR (corrected 95% CI)	Corrected p value	OR (corrected 95% CI)	Corrected p value	OR (corrected 95% CI)	Corrected p value	OR (corrected 95% CI)	Corrected p value
Major depressive disorder, F32	2·14 (1·77–2·59)	<0·0001	3·41 (3·04–3·84)	<0·0001	1·34 (1·15–1·56)	<0·0001	1·73 (1·45–2·07)	<0·0001	1·13 (0·87–1·45)	1·0	1·07 (0·75–1·53)	1·0
Anxiety, F41	2·02 (1·69–2·41)	<0·0001	1·93 (1·71–2·17)	<0·0001	1·36 (1·18–1·56)	<0·0001	1·5 (1·26–1·78)	<0·0001	1·07 (0·85–1·35)	1·0	1·05 (0·73–1·52)	1·0
Reaction to severe stress and adjustment disorders, F43	2·34 (1·76–3·11)	<0·0001	2·1 (1·61–2·75)	<0·0001	1·4 (1·12–1·77)	<0·0001	1·83 (1·36–2·46)	<0·0001	0·94 (0·58–1·52)	1·0	1·48 (0·83–2·63)	1·0
Hearing loss, H91	1·28 (1·07–1·52)	0·0001	1·95 (1·35–2·82)	<0·0001	1·19 (1·04–1·36)	0·0006	1·51 (1·01–2·26)	0·033	1·11 (0·85–1·44)	1·0	..*	..*
Constipation, K59	1·41 (1·23–1·63)	<0·0001	1·66 (1·47–1·87)	<0·0001	1·31 (1·16–1·49)	<0·0001	1·59 (1·33–1·89)	<0·0001	1·18 (0·92–1·51)	1·0	1·35 (0·92–1·99)	0·75
Spondylosis, M47	1·25 (0·93–1·69)	0·933	1·45 (1·22–1·72)	<0·0001	1·26 (1·05–1·5)	0·0005	1·62 (1·32–1·98)	<0·0001	1·12 (0·85–1·47)	1·0	1·21 (0·81–1·83)	1·0
Memory loss symptom, R41	31·5 (24·18–41·05)	<0·0001	16·5 (10·39–26·19)	<0·0001	7·63 (5·95–9·79)	<0·0001	4·41 (2·3–8·48)	<0·0001	1·59 (0·72–3·54)	1·0	..*	..*
Malaise and fatigue, R53	1·36 (1·17–1·58)	<0·0001	1·78 (1·59–2·0)	<0·0001	1·23 (1·08–1·39)	<0·0001	1·59 (1·36–1·86)	<0·0001	1·11 (0·88–1·4)	1·0	1·36 (0·96–1·93)	0·25
Syncope and collapse, R55	1·95 (1·53–2·48)	<0·0001	2·49 (1·68–3·69)	<0·0001	1·23 (1·01–1·5)	0·034	1·57 (1·06–2·34)	0·007	1·22 (0·82–1·82)	1·0	..*	..*
Abnormal weight loss, R63	2·1 (1·68–2·62)	<0·0001	3·12 (2·41–4·02)	<0·0001	1·47 (1·22–1·77)	<0·0001	1·88 (1·35–2·62)	<0·0001	1·1 (0·73–1·66)	1·0	1·89 (0·67–5·28)	1·0

Both CIs and p values were corrected for multiple comparisons. OR=odds ratio. *Cannot be calculated because an insufficient number of presentations was recorded.

Table 2: **ORs for all variables individually associated with a future diagnosis of Alzheimer's disease in the 2–10 years before diagnosis**

date in the UK, whereas in France, patients with Alzheimer's disease with at least 10 years of retrospective data before diagnosis had a median age of 82 years (77–86) at the index date (versus 80 years [75–85] for the full analytical cohort). In the UK study, 514 (2·6%) of 19 940 patients had a diagnosis of non-Alzheimer's disease dementia before the index date, with ICD10 codes F01 for vascular dementia, F02 for dementia in other diseases classified elsewhere, F03 for unspecified dementia, and G31.1 for senile degeneration of the brain; 148 patients had a diagnosis of non-Alzheimer's disease dementia with one of these codes at least 2 years before index date. In the French study, 185 (1·0%) of 19 056 patients had one of these codes recorded before the index date and 65 (0·3%) patients had one of these codes recorded at least 2 years before the index date. In the control groups, 448 (2·2%) of 19 940 individuals in the UK study and 172 (0·9%) of 19 056 individuals in the French study had a diagnosis of non-Alzheimer's disease dementia recorded before the index date (268 [1·3%] individuals and 89 [0·5%] individuals at least 2 years before the index date, respectively). The mean number of visits per year was similar for the case and control groups in both countries. In the control groups, for both countries, the health condition most frequently recorded in the databases was essential hypertension, with a prevalence of 9831 (48·6%) individuals in the UK and 11 070 (56·9%) individuals in France, followed by cough (8494 [42·0%]) and dorsalgia (8403 [41·6%]) in the UK and by dorsalgia (7374 [37·9%]) and hypercholesterolaemia (6224 [32·0%]) in France (appendix p 1). The prevalence of type 2 diabetes in the control group was 2574 (12·7%) in the UK and 2606 (13·4%) in France.

We agnostically tested 123 possible health conditions (0–2] years, (2–10] years, and (10–15] years before the diagnosis of Alzheimer's disease (appendix p 8). Only ten (8%) of the 123 health conditions were significantly positively associated with the diagnosis of Alzheimer's disease in both countries after Bonferroni correction in the (2–10]-year time window (table 2; figure 1). None of these health conditions were significantly associated with the diagnosis of Alzheimer's disease in the (10–15]-year time frame in which no health condition was found to increase the risk of Alzheimer's disease onset in both countries. Spondylosis was the only health condition significantly associated with Alzheimer's disease diagnosis in the (2–10]-year period but not in the (0–2]-year period, this relationship being observed only in the UK. Cervical spondylosis accounted for 63% of the reported spondylosis cases in the UK and 60% in France. Another 12 health conditions were significantly positively associated with Alzheimer's disease in both countries in the (0–2]-year time frame (appendix p 8). Five of the ten health conditions identified as being significantly associated with Alzheimer's disease (anxiety, constipation, spondylosis,

memory loss, and abnormal weight loss) remained significant (p<0·050) in the conditional logistic model containing all health conditions in the UK (table 3). They were also significant in the French study (table 3). Assuming causality for depression, anxiety, spondylosis, constipation, hearing loss, and reaction to severe stress, the potential attributable risk associated with these health conditions was 14% in France and 18% in the UK.

Finally, we showed the change in incidence over time in the control and Alzheimer's disease cohorts, for the health conditions significantly associated with Alzheimer's disease risk (figures 2 and 3). We found that depression and anxiety were the first health conditions to be associated with Alzheimer's disease, at least 9 years before the clinical diagnosis of Alzheimer's disease, followed by constipation and abnormal weight loss 7 years before the index date. These findings facilitate differentiation between risk factors associated with early stages of the disease and probable comorbid conditions occurring within a few years of diagnosis (eg, sleep disorders and urinary system disorders).

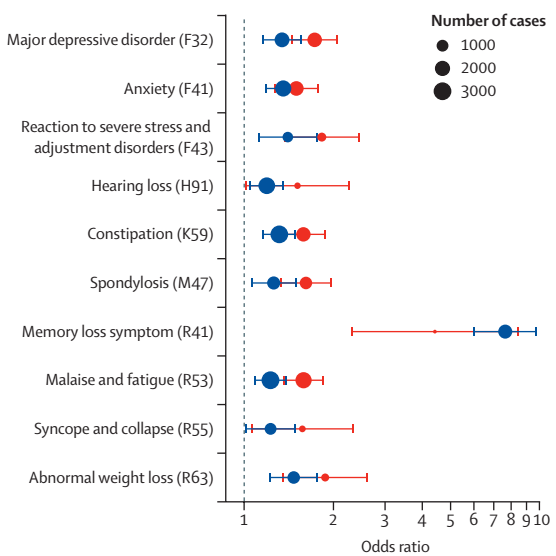

Figure 1: **Health conditions associated with a future diagnosis of Alzheimer's disease in the 2–10 years before diagnosis**
Only risk factors significant in both the French and UK cohorts are shown, with orange and blue dots, respectively. The size of the dot is proportional to the number of affected people. Bars correspond to 95% CIs after correction for multiple comparisons.

Discussion

This large agnostic study of patients from two European countries showed that several health conditions observed by general practitioners are significantly associated with the diagnosis of Alzheimer's disease several years later. In these two studies, which included a total of 39 672 patients with Alzheimer's disease, ten of the 123 health conditions considered were significantly associated with the subsequent diagnosis of Alzheimer's disease in both countries and could be considered potential prediagnostic features. Depression and anxiety were among the risk factors for dementia identified by the recent *Lancet* Commission.[6] Nevertheless, it remains a matter of debate whether these conditions are risk factors for dementia, early symptoms of dementia, or both.[4] The association between depression and subsequent Alzheimer's disease diagnosis became significant at least 9 years before the first clinical diagnosis of Alzheimer's disease, consistent with the Whitehall II cohort study report.[4]

Hearing loss[20,21] has also been identified as a potential risk factor for dementia in several observational and prospective studies.[22–24] A recent genetic study[25] showed a correlation between genetically determined hearing impairment and Alzheimer's disease. However, no causal link could be demonstrated between these traits, to confirm or deny the hypothesis that the simple

	Univariable analysis				Multivariable analysis			
	UK		France		UK		France	
	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value
Major depressive disorder, F32	1·34 (1·23–1·46)	<0·0001	1·73 (1·57–1·91)	<0·0001	1·07 (0·97–1·17)	0·175	1·41 (1·27–1·57)	<0·0001
Anxiety, F41	1·36 (1·25–1·47)	<0·0001	1·50 (1·36–1·65)	<0·0001	1·16 (1·07–1·27)	0·001	1·16 (1·04–1·28)	0·006
Reaction to severe stress and adjustment disorders, F43	1·40 (1·24–1·59)	<0·0001	1·83 (1·55–2·15)	<0·0001	1·03 (0·89–1·18)	0·721	1·38 (1·16–1·64)	0·0003
Hearing loss, H91	1·19 (1·11–1·28)	<0·0001	1·51 (1·21–1·89)	0·0003	1·08 (1·00–1·17)	0·057	1·28 (1·01–1·61)	0·039
Constipation, K59	1·31 (1·22–1·41)	<0·0001	1·59 (1·44–1·75)	<0·0001	1·19 (1·10–1·28)	<0·0001	1·28 (1·16–1·42)	<0·0001
Spondylosis, M47	1·26 (1·14–1·39)	<0·0001	1·62 (1·44–1·81)	<0·0001	1·16 (1·05–1·29)	0·004	1·33 (1·18–1·50)	<0·0001
Memory loss, R41	7·63 (6·65–8·76)	<0·0001	4·41 (3·07–6·34)	<0·0001	7·28 (6·33–8·36)	<0·0001	3·63 (2·51–5·26)	<0·0001
Malaise and fatigue, R53	1·23 (1·14–1·32)	<0·0001	1·59 (1·46–1·73)	<0·0001	1·01 (0·93–1·09)	0·809	1·25 (1·14–1·37)	<0·0001
Syncope and collapse, R55	1·23 (1·10–1·37)	0·0003	1·57 (1·26–1·96)	<0·0001	1·07 (0·95–1·20)	0·292	1·23 (0·98–1·55)	0·078
Abnormal weight loss, R63	1·47 (1·33–1·63)	<0·0001	1·88 (1·56–2·26)	<0·0001	1·25 (1·12–1·40)	<0·0001	1·39 (1·14–1·68)	0·001

In the multivariable model, we included all ten health conditions. These health conditions were measured 2–10 years before Alzheimer's disease diagnosis. OR=odds ratio.

Table 3: **OR and 95% CIs for a conditional logistic model assessing the effect of all variables individually associated with a future diagnosis of Alzheimer's disease in the UK**

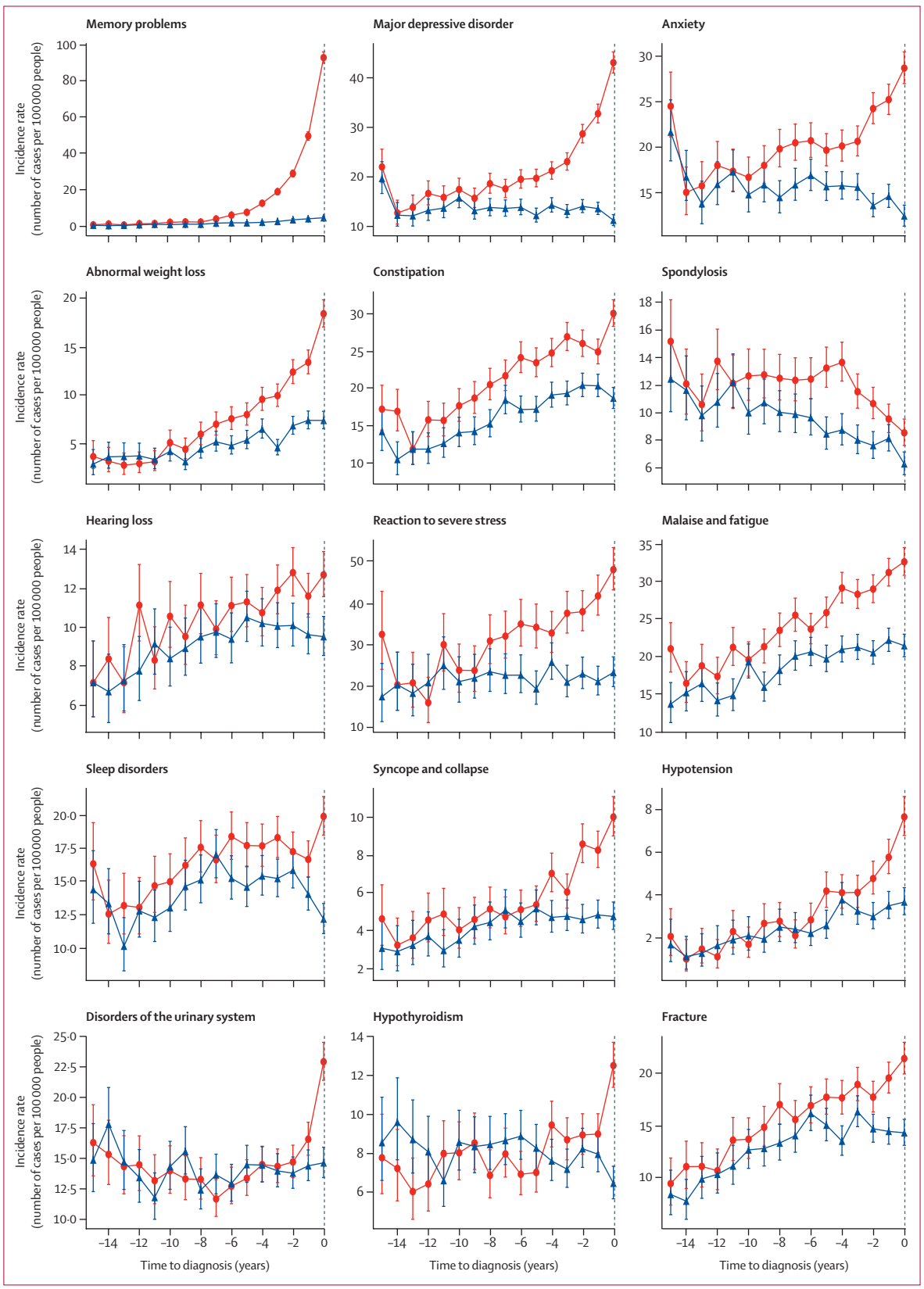

***Figure 2*: Incidence of comorbid conditions significantly associated with a subsequent Alzheimer's disease diagnosis**
The red curve corresponds to patients with Alzheimer's disease and the blue curve corresponds to controls. Bars indicate 95% CIs, on the basis of the sample size of each group.

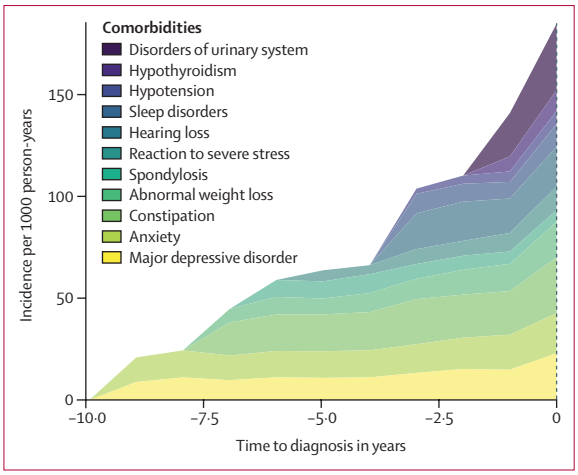

Figure 3: **Change in incidence rate over time in the Alzheimer's disease cohort for health conditions significantly associated with the disease**
The incidence rate is in number of cases per 1000 person-years.

management of hearing loss can decrease the risk of Alzheimer's disease. The association between Alzheimer's disease and hearing loss did not remain significant in the UK after adjustment for other risk factors as possible confounders (p=0·057). We also confirmed associations between Alzheimer's disease and abnormal weight loss,[7] which remained significant in the multivariate model.

The agnostic approach used here made it possible to identify other potential new risk factors for Alzheimer's disease. Constipation has already been linked to depression[26] and has been identified as a symptom in some neurodegenerative diseases, such as Lewy-body dementia and Parkinson's disease.[14] To our knowledge, this is the first time that constipation has been shown to be associated with a risk of Alzheimer's disease. This association was detected 7 years before the first clinical diagnosis of Alzheimer's disease, and remained significant until the time of Alzheimer's disease diagnosis and after controlling for depression and anxiety. We also found a strong association between spondylosis and Alzheimer's disease risk, which remained significant after correcting for other identified risk factors. One possible reason for this association might be a decline in physical activity[6] caused by spondylosis. A higher prevalence of dementia and Alzheimer's disease was reported in a retrospective longitudinal study of patients with ankylosing spondylitis.[27] Nevertheless, we found an association with spondylosis (ICD10 code M47), but not with ankylosing spondylitis (ICD10 code M49). This result suggests that the association might not be caused by the inflammatory nature of spondylitis. We noted that the spondylosis was cervical in most cases. Therefore, investigating whether cervical spondylosis affects blood flow[28] or cerebrospinal fluid[29] to the brain, thereby favouring or accelerating neurodegeneration, might be worthwhile. The associations of Alzheimer's disease with syncope and hypotension were also documented in a recent study reporting an association between midlife orthostatic hypotension and future dementia.[17]

We found associations present only in the last 2 years before diagnosis. The lateness of these associations suggests that these conditions are probably prodromal symptoms of the disease rather than risk factors. For instance, a Mendelian randomised study[18] showed that Alzheimer's disease was more likely to be the cause of sleep disorders than a consequence of these disorders. The association with hypothyroidism might, therefore, reflect active investigations in patients with cognitive impairment,[19] because the evaluation of thyroid hormone concentrations is recommended for this indication by WHO.

We also highlighted risk factors that are not replicated in the analysis. Hypertension has repeatedly been presented as a potential modifiable risk factor for Alzheimer's disease,[30] whereas our data suggest a protective effect in the UK study and no association in the French study (France OR 1·01, 95% CI 0·87–1·18; UK OR 0·82, 0·75–0·90). As in other observational studies,[31,32] this association might result from a channelling effect, in which patients with dementia and hypertension or other cardiovascular comorbidities are more likely to be diagnosed with vascular dementia than Alzheimer's disease.

We did not observe an association between type 2 diabetes or other types of diabetes and Alzheimer's disease, although associations between diabetes and dementia have been reported.[33] The prevalence of type 2 diabetes in the control groups is 12·7% in the UK and 13·4% in France, which is compatible with what was reported in previous studies.[14]

Other previously reported associations not replicated in this study included associations with herpes virus infections,[34] alcohol use disorders,[35] and obesity.[36] More work is needed to fully understand the relationship between these potential risk factors and dementia subtypes.[33]

Epidemiological approaches, such as *Lancet* Commissions,[3,6] tend to focus on all-cause dementia, whereas other approaches in dementia research tend to argue for increasingly precise biological definitions of Alzheimer's disease.[37] The early detection and labelling of biomarker-positive patients before the onset of symptoms has been criticised, due to the uncertainty of progression to dementia.[38] This study has the advantage of bridging the gap between these different approaches, through a large-scale, exploratory, data-driven approach, considering all the main health conditions observable at the point of care.

Beyond its data-driven precision, another notable strength of this study is its inclusion of sufficient numbers of patients from two countries, making it possible to replicate the analysis, with consistent results obtained for the two samples, ensuring a high degree of generalisability.

We found that the results generalised well between France and the UK. Both health systems are national and

public. The question remains open whether the results would generalise also in other health systems where accessibility might be more limited due to cost or opportunities. Another question raised by this study is whether the variables identified are genuine risk factors, which could be tested in Mendelian randomisation studies on databases including genetic data[18] or in new prospective research cohorts.

The study also has some limitations. First, we did not have data about potential confounding factors such as education level, ethnicity, socioeconomic status, body-mass index, or genetic information. To assess the extent that these unmeasured confounders could have led to biased estimates, we computed e values that represented the minimum strength of association that an unmeasured confounder would need to have with both the exposure and the outcome to disregard a specific exposure–outcome association.[39] In the French study, the smallest OR was observed for anxiety (OR 1·50, 95% CI 1·36–1·65). The corresponding e value was 1·75, meaning that an unmeasured confounding factor would need to be associated with a minimum OR of 1·75 with both anxiety and Alzheimer's disease dementia to disregard the anxiety and Alzheimer's disease dementia association.

Second, our findings were based on primary care data. No hospital admission or mortality records were used so severe and complicated forms of Alzheimer's disease might have been missed. However, these forms of the disease are rare and unlikely to affect our results. Third, controls were selected from primary care databases, and therefore did not include individuals who never consulted a general practitioner. Controls were also not allowed to have records of the listed neurodegenerative diseases throughout their whole observable primary care records. Nevertheless, they do include other types of dementia, such as vascular dementia. Finally, the diagnosis of Alzheimer's disease was made by a general practitioner, which might result in a later diagnosis than in prospective cohorts in which each patient follows a protocol with planned visits with a neurologist. This is the main reason why we tested health conditions recorded at least 2 years before the Alzheimer's disease diagnosis.

This study provides a modern extensive data-driven survey of potential risk factors and early symptomatic expression for Alzheimer's disease in primary care, and constitutes a first attempt to determine when these comorbid conditions appear during the course of the disease. Our findings make it possible to model the possible trajectories of risk factors in the period preceding the diagnosis of Alzheimer's disease, providing new insights into possible windows for prevention.

Contributors

TN, BC-D, SE, CD, and SD conceived and designed the study and analysed the data. FM, MA, LG, and BL extracted the data. FM, MA, and TN ensured data quality. TN and BC-D generated the figures. All authors interpreted the results and contributed to the writing of the final version of the manuscript. Editorial support, in the form of substantive editing, was provided by Julie Sappa of Alex Edelman and Associates.

This assistance was funded by internal laboratory resources. All authors agreed with the results and conclusions and approved the final draft. All authors had full access to all the data in the study and the corresponding author had final responsibility for the decision to submit for publication.

Declaration of interests

SE reports personal fees from Biogen, Eisai, Roche, and GE Healthcare, for presentations or participation on advisory boards. FM, LG, and BE are full time employees of Cegedim. All other authors declare no competing interests.

Data sharing

The data used in the preparation of this Article are available from the Cegedim company upon reasonable request (info@the-health-improvement-network.co.uk).

Acknowledgments

This work was funded in part by the Agence Nationale de la Recherche, as part of the Investissements d'Avenir programme, under grants ANR-10-IAIHU-06 (Institut Hospitalo-Universitaire-Institut du Cerveau et de la Moelle) and ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). BC-D is supported by a CJ Martin Fellowship (National Health and Medical Research Council APP1161356).

References

1 Alzheimer's Association. 2021 Alzheimer's disease facts and figures. *Alzheimers Dement* 2021; **17:** 327–406.
2 Niu H, Álvarez-Álvarez I, Guillén-Grima F, Aguinaga-Ontoso I. Prevalence and incidence of Alzheimer's disease in Europe: a meta-analysis. *Neurología* 2017; **32:** 523–32.
3 Livingston G, Sommerlad A, Orgeta V, et al. The *Lancet* International Commission on Dementia Prevention and Care. *Lancet* 2017; **390:** 2673–734.
4 Singh-Manoux A, Dugravot A, Fournier A, et al. Trajectories of depressive symptoms before diagnosis of dementia: a 28-year follow-up study. *JAMA Psychiatr* 2017; **74:** 712–18.
5 Frisoni GB, Molinuevo JL, Altomare D, et al. Precision prevention of Alzheimer's and other dementias: anticipating future needs in the control of risk factors and implementation of disease-modifying therapies. *Alzheimers Dement* 2020; **16:** 1457–68.
6 Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the *Lancet* Commission. *Lancet* 2020; **396:** 413–46.
7 Yu J-T, Xu W, Tan C-C, et al. Evidence-based prevention of Alzheimer's disease: systematic review and meta-analysis of 243 observational prospective studies and 153 randomised controlled trials. *J Neurol Neurosurger Psychiatr* 2020; **91:** 1201–09.
8 Muzambi R, Bhaskaran K, Brayne C, Davidson JA, Smeeth L, Warren-Gash C. Common bacterial infections and risk of dementia or cognitive decline: a systematic review. *J Alzheimers Dis* 2020; **76:** 1609–26.
9 Sipilä PN, Heikkilä N, Lindbohm JV, et al. Hospital-treated infectious diseases and the risk of dementia: a large, multicohort, observational study with a replication cohort. *Lancet Infect Dis* 2021; **21:** 1557–67.
10 Sipilä PN, Lindbohm JV, Singh-Manoux A, et al. Long-term risk of dementia following hospitalization due to physical diseases: a multicohort study. *Alzheimers Dement* 2020; **16:** 1686–95.
11 Ford E, Greenslade N, Paudyal P, et al. Predicting dementia from primary care records: a systematic review and meta-analysis. *PLoS One* 2018; **13:** e0194735.
12 Blak B, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *J Innov Health Inform* 2011; **19:** 251–55.
13 Iordache L, Strazzulla A, Lekens B, Bergmann J, Diamantis S. Évaluation du suivi des recommandations de prise en charge des infections urinaires en médecine de ville en France. *Med Malad Infect* 2018; **48:** S101–02.
14 Schrag A, Horsfall L, Walters K, Noyce A, Petersen I. Prediagnostic presentations of Parkinson's disease in primary care: a case-control study. *Lancet Neurol* 2015; **14:** 57–64.
15 Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Brit J Gen Pract* 2010; **60:** e128–36.

16 Donnelly K. SNOMED-CT: the advanced terminology and coding system for eHealth. *Stud Health Technol Inform* 2006; **121:** 279.

17 Rawlings AM, Juraschek SP, Heiss G, et al. Association of orthostatic hypotension with incident dementia, stroke, and cognitive decline. *Neurology* 2018; **91:** e759–68.

18 Huang J, Zuber V, Matthews PM, Elliott P, Tzoulaki J, Dehghan A. Sleep, major depressive disorder, and Alzheimer disease: a Mendelian randomization study. *Neurology* 2020; **95:** e1963–70.

19 Pasqualetti G, Pagano G, Rengo G, Ferrara N, Monzani F. Subclinical hypothyroidism and cognitive impairment: systematic review and meta-analysis. *J Clin Endocrinol Metabol* 2015; **100:** 4240–48.

20 Amieva H, Ouvrard C. Does treating hearing loss in older adults improve cognitive outcomes? A review. *J Clin Med* 2020; **9:** 805.

21 Panza F, Lozupone M, Sardone R, et al. Sensorial frailty: age-related hearing loss and the risk of cognitive impairment and dementia in later life. *Ther Adv Chronic Dis* 2019; **10:** 2040622318811000.

22 Amieva H, Ouvrard C, Giulioli C, Meillon C, Rullier L, Dartigues J-F. Self-reported hearing loss, hearing aids, and cognitive decline in elderly adults: a 25-year study. *J Am Geriat Soc* 2015; **63:** 2099–104.

23 Deal JA, Betz J, Yaffe K, et al. Hearing impairment and incident dementia and cognitive decline in older adults: the health ABC study. *J Gerontol A Biol Sci Med Sci* 2017; **72:** 703–09.

24 Gallacher J, Ilubaera V, Ben-Shlomo Y, et al. Auditory threshold, phonologic demand, and incident dementia. *Neurology* 2012; **79:** 1583–90.

25 Mitchell BL, Thorp JG, Evans DM, Nyholt DR, Martin NG, Lupton MK. Exploring the genetic relationship between hearing impairment and Alzheimer's disease. *Alzheimers Dement* 2020; **12:** e12108.

26 Del Colle A, Israelyan N, Li Z, et al. 1085–neuronal serotonin as a link between depression and constipation: a role for simultaneous treatment with a novel slow-release formulation of the serotonin precursor, 5-hydroxytryptophan. *Gastroenterol* 2019; **156:** S229.

27 Jang H-D, Park J-S, Kim DW, et al. Relationship between dementia and ankylosing spondylitis: a nationwide, population-based, retrospective longitudinal cohort study. *PLoS One* 2019; **14:** e0210335.

28 Roher AE, Debbins JP, Malek-Ahmadi M, et al. Cerebral blood flow in Alzheimer's disease. *Vasc Health Risk Manag* 2012; **8:** 599.

29 Rubenstein E. Relationship of senescence of cerebrospinal fluid circulatory system to dementias of the aged. *Lancet* 1998; **351:** 283–85.

30 Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol* 2014; **13:** 788–94.

31 Bohlken J, Jacob L, Kostev K. The relationship between the use of antihypertensive drugs and the incidence of dementia in general practices in Germany. *J Alzheimers Dis* 2019; **70:** 91–97.

32 Imfeld P, Brauchli Pernus YB, Jick SS, Meier CR. Epidemiology, co-morbidities, and medication use of patients with Alzheimer's disease or vascular dementia in the UK. *J Alzheimers Dis* 2013; **35:** 565–73.

33 Benn M, Nordestgaard BG, Tybjærg-Hansen A, Frikke-Schmidt R. Impact of glucose on risk of dementia: Mendelian randomisation studies in 115 875 individuals. *Diabetologia* 2020; **63:** 1151–61.

34 Harris SA, Harris EA. Molecular mechanisms for herpes simplex virus type 1 pathogenesis in Alzheimer's disease. *Front Aging Neuro* 2018; **10:** 48.

35 Schwarzinger M, Pollock BG, Hasan OS, et al. Contribution of alcohol use disorders to the burden of dementia in France 2008–13: a nationwide retrospective cohort study. *Lancet Public Health* 2018; **3:** e124–32.

36 Whitmer R, Gustafson D, Barrett-Connor E, Haan M, Gunderson E, Yaffe K. Central obesity and increased risk of dementia more than three decades later. *Neurology* 2008; **71:** 1057–64.

37 Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 2018; **14:** 535–62.

38 Schermer MH, Richard E. On the reconceptualization of Alzheimer's disease. *Bioethics* 2019; **33:** 138–45.

39 VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Int Med* 2017; **167:** 268–74.