

Machine learning-assisted optimization of dietary intervention against dementia risk

Received: 5 June 2024

Accepted: 27 May 2025

Published online: 02 July 2025

 Check for updates

Si-Jia Chen^{1,11}, Hui Chen^{1,21}, Jia You^{1,3,11}, Shi-Dong Chen^{1,11}, Yan Fu⁴, Wei Zhang³, Liyan Huang², Jian-Feng Feng^{1,3,5,6,7,8}, Xiang Gao⁹, Wei Cheng^{1,3,5,6}, Changzheng Yuan^{1,2,10}✉ & Jin-Tai Yu¹✉

A healthy diet has been associated with a reduced risk of dementia. Here we devised a Machine learning-assisted Optimizing Dietary intERvention against demeNtia risk (MODERN) diet based on data from 185,012 UK Biobank participants, 1,987 of whom developed all-cause dementia over 10 years. We first identified 25 food groups associated with dementia in a food-wide association analysis. Second, we ranked their importance using machine learning and prioritized eight groups (for example, green leafy vegetables, berries and citrus fruits). Finally, we established and externally validated a MODERN score (0–7), which showed stronger associations with lower risk of dementia-related outcomes (hazard ratio comparing highest versus lowest tertiles: 0.64, 95% CI: 0.43–0.93) than the a priori-defined MIND diet (0.75, 0.61–0.92). Across 63 health-related outcomes, the MODERN diet showed particularly significant associations with mental/behavioural disorders. Multimodal neuroimaging, metabolomics, inflammation and proteomics analyses revealed potential pathways and further support the potential of MODERN diet for dementia prevention.

Dementia poses substantial societal and healthcare challenges to the worldwide aging population¹. The lack of effective therapy makes dementia prevention of great importance. Among a series of potentially modifiable factors of dementia, dietary factors raise major interest, but evidence in this area is still limited and inconsistent^{2,3}. Furthermore, while observational studies imply that a healthy diet may delay dementia onset^{4–6} and slow brain structural and pathological changes^{7–9}, there is a pressing need to identify dietary patterns tailored for brain health that are both effective and practical for implementation in intervention strategies.

Among multiple healthy dietary patterns, the Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) diet, devised

in 2015 for the US population with a focus on brain health, recommends intake of ten food groups and restricts five (DASH stands for Dietary Approaches to Stop Hypertension). While observational studies have shown its potential in slowing cognitive decline and delaying dementia onset^{2,10–12}, a 3-year randomized controlled trial (RCT)¹³ did not find its additional cognitive benefits beyond a calorie-restricted diet. Existing observational studies are often hampered by relatively small sample sizes, short-term follow-up, or both, especially considering the long preclinical phase of dementia. Although RCTs are considered the gold standard for causal inference, their implementation in long-term dietary interventions is hindered by challenges such as

¹Department of Neurology and National Center for Neurological Disorders, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, Shanghai, China. ²School of Public Health, the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ³Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. ⁴Department of Neurology, Qingdao Municipal Hospital, Qingdao University, Qingdao, China. ⁵Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China. ⁶Fudan ISTBI—ZJNU Algorithm Centre for Brain-Inspired Intelligence, Zhejiang Normal University, Jinhua, China. ⁷MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. ⁸Zhangjiang Fudan International Innovation Center, Shanghai, China. ⁹Department of Nutrition and Food Hygiene, School of Public Health, Institute of Nutrition, Fudan University, Shanghai, China. ¹⁰Department of Nutrition, Harvard T H Chan School of Public Health, Boston, MA, USA. ¹¹These authors contributed equally: Si-Jia Chen, Hui Chen, Jia You, Shi-Dong Chen. ✉e-mail: chy478@zju.edu.cn; jintai_yu@fudan.edu.cn

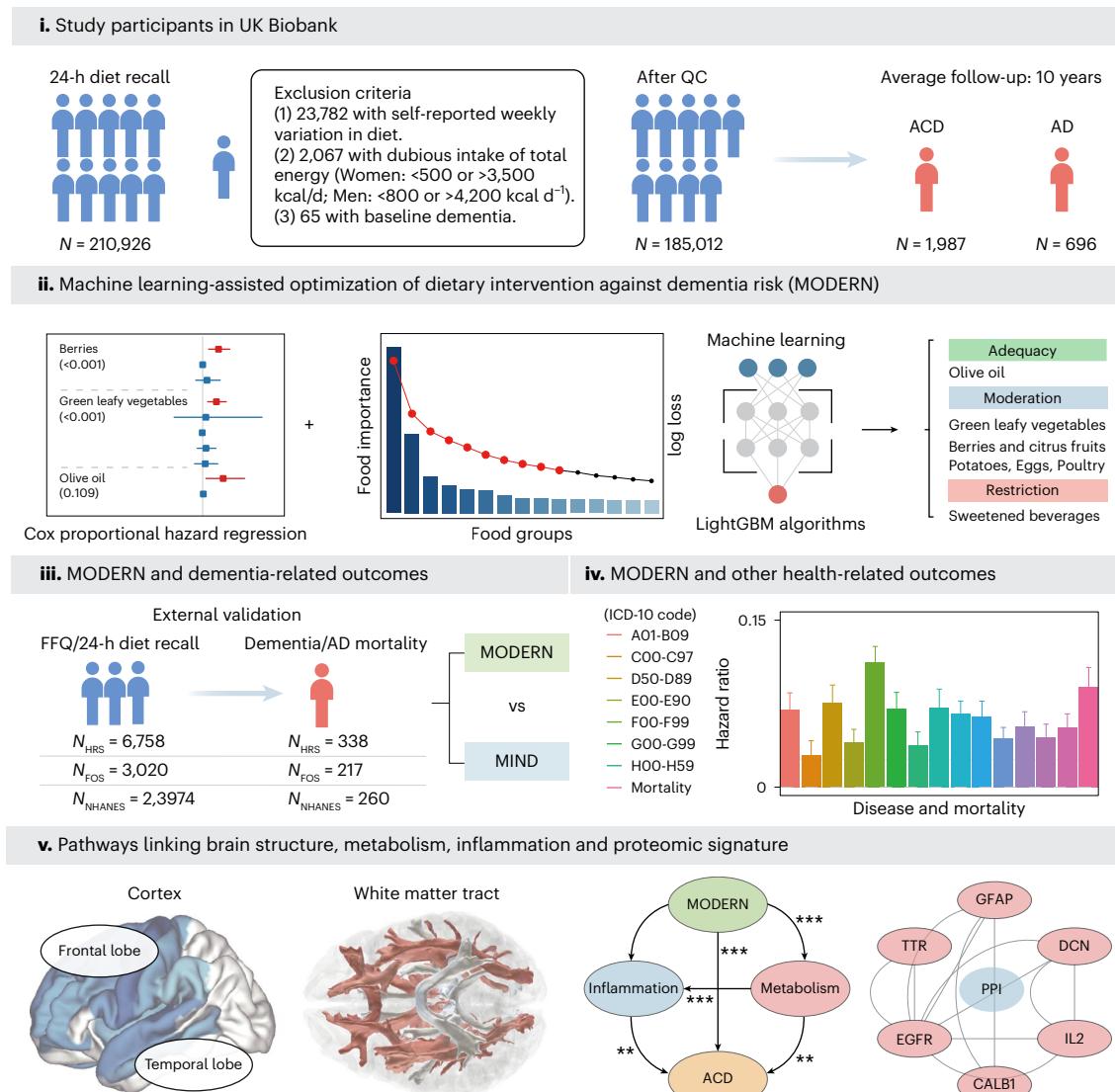


Fig. 1 | Design of the study. The study included a total of 185,012 participants with 24-h diet recalls. After an average follow-up of 10 years, 1,987 participants developed ACD. First, food-wide association analyses of 34 food groups of 206 foods from the Oxford WebQ with incident ACD were conducted using multivariable Cox proportional hazard regression models. Second, by machine learning, eight food groups were selected to construct a new diet pattern called

MODERN, with seven components. MODERN's association with incident ACD was first examined in UKB and then validated in three external cohorts: HRS, FOS and NHANES. Furthermore, its associations with other health-related outcomes were examined in UKB. Finally, the potential pathways linking brain structure, metabolism, inflammation and proteomic signature were investigated. QC, quality control.

maintaining participant adherence and sustained follow-up. Therefore, simpler and more practical dietary patterns are warranted to enable large-scale intervention for healthy brain aging. With the accumulation of evidence in the diet–cognition field, the a priori-defined MIND diet may be further refined to incorporate potential linear and nonlinear associations between food groups and dementia in a larger long-term cohort study and necessitate population-specific redesign and modifications.

Motivated by the aforementioned research gaps, the present study aimed to devise a dietary pattern for dementia prevention by combining a large-scale cohort study with a machine-learning approach. Leveraging data from the UK Biobank (UKB) with 185,012 participants, we first conducted a food-wide association analysis and identified food groups associated with incident dementia. We then ranked the importance of food groups in predicting dementia risk with a machine-learning approach and derived a new dietary pattern for dementia prevention. Finally, we compared the magnitude of its association with dementia with the MIND diet in external validation

settings and explored the potential mechanisms using multimodal and multi-omics data (Fig. 1).

Results

Food-wide association study with incident all-cause dementia

The present study included 185,012 participants (mean age [s.d.]: 59.2 [7.97] years; women: 55.3%) with 24-h diet recall data from the UKB (Supplementary Table 1). Thirty-four food groups were amalgamated from 206 food items across 10 categories assessed by the Oxford WebQ, a validated tool for 24-h diet recall¹⁴ (Supplementary Table 2). During an average follow-up of 10 years, 1,987 participants developed all-cause dementia (ACD), including 696 with Alzheimer's disease (AD). The mean age of onset was 74.6 years for ACD and 75.0 years for AD (Fig. 1). In the fully adjusted Cox proportional hazard model, intake levels of 25 food groups were associated with the risk of ACD (Fig. 2 and Supplementary Table 3). The associations were more pronounced for citrus fruits (hazard ratio [HR]_{3 levels}|Level3vsLevel2: 1.26; 95% confidence interval [CI]: 1.09–1.45), other vegetables (HR_{SL|LvsL4}: 1.38; 95% CI: 1.21–1.57), refined

grains ($HR_{SL|LvsL3}$: 1.31; 95% CI: 1.10–1.56), processed meats ($HR_{4L|LvsL2}$: 1.41; 95% CI: 1.16–1.72), eggs ($HR_{3L|LvsL2}$: 1.27; 95% CI: 1.12–1.45), olive oil ($HR_{2L|LvsL2}$: 1.28; 95% CI: 1.04–1.59) and sweetened beverages ($HR_{4L|LvsL1}$: 1.38; 95% CI: 1.22–1.56) (Fig. 2). The restricted cubic spline (RCS) analyses revealed 'U'-shaped relationships for multiple food groups other than other fruits, olive oil and sweetened beverages (Fig. 2 and Supplementary Fig. 1). Sweetened beverage was associated with higher ACD risk in a nearly dose-response manner. The above associations remained almost unchanged after excluding participants with major diseases at baseline, including diabetes, hypertension, cerebrovascular diseases and other cardiovascular diseases (CVDs) (Supplementary Table 4). Furthermore, in each food group, some of the constituent foods showed significant associations with ACD risk after false discovery rate (FDR) correction (Supplementary Fig. 2 and Table 5). Specifically, in citrus fruits, intake (versus non-intake) of grapefruit was associated with a lower risk of ACD (HR : 0.92; P_{FDR} : 0.009). Similarly, intake of sweet pepper and tomato (in red/orange vegetables) was associated with a lower risk (HR : 0.90–0.91; P_{FDR} : 0.001–0.004). In sweetened beverages, intake of dairy smoothies was associated with a higher risk of ACD (HR : 1.07; P_{FDR} : 0.006).

Machine learning-assisted construction of MODERN diet score
For 25 food groups individually associated with ACD, we primarily used a machine-learning approach (LightGBM) to identify the optimal combination for dementia prevention. With a forward selection scheme, ten food groups were identified according to their importance in predicting dementia risk, including poultry, potatoes, eggs, citrus fruits, sweetened beverages, olive oil, berries, beer, water and green leafy vegetables (Fig. 3a). Additional experiments with various machine-learning-based approaches yielded consistent results (Methods), and poultry, potatoes, eggs, citrus fruits, sweetened beverages, olive oil and berries were selected by at least six experiments (Supplementary Table 6). SHapley Additive exPlanations (SHAP) analysis demonstrated strong concordance in model-specific feature importance rankings, with substantial overlap in the top-ranked food groups across LightGBM (9/10), XGBoost (8/10) and Random Forest (7/10) (Supplementary Table 6).

Predictive analyses were performed to evaluate the discriminative capacity of selected food groups. Among multiple candidate machine-learning models using the top-10 food groups, decision tree-based algorithms demonstrated similarly higher performance, with area under the receiver operating characteristic (ROC) curve (AUC) values ranging from 0.698 to 0.722. Specifically, LightGBM achieved an AUC of 0.722 (95% CI: 0.711–0.732) using model-based food group selection, demonstrating superior predictive performance (highest AUC) in comparative methods. When using the full set of 22 food groups, LightGBM's performance slightly decreased (AUC: 0.717), suggesting that feature selection may enhance model performance (Supplementary Table 7).

We devised a dietary pattern consisting of seven components, named Machine learning-assisted Optimizing Dietary intErvention against demeNtia risk (MODERN), on the basis of eight of these food groups (except for beer and water) (Fig. 3b). A composite diet score was then constructed by rating the intake levels from the associations of each component with risk of ACD, as elaborated in Methods (Supplementary Table 3 and Fig. 1). Briefly, the score for each component ranged from 0 to 1 and the total MODERN score ranged from 0 to 7 (highest adherence). Olive oil (average intake level >0 servings per day) was categorized as adequacy components and recommended for higher intake. Green leafy vegetables (0–1.5 servings per day), berries and citrus fruits (0–2 servings per day), potatoes (0–0.75 servings per day), eggs (0–1 servings per day) and poultry (0–0.5 servings per day) were categorized as moderation components and recommended for moderate intake. Sweetened beverages were categorized as restriction components (0 servings per day; Fig. 3b). Each component score was significantly associated with a lower ACD risk (HR_{Lvs0} : 0.78–0.84; Extended Data Fig. 1 and Supplementary Table 8).

MODERN and MIND diet scores and incident dementia risk

In UKB, participants in the higher quartile of the MODERN diet score were more likely to have higher total energy intake, be older, women, white, have lower Townsend deprivation index (TDI), higher education, be non-smokers, have higher physical activity, lower body mass index (BMI), higher fluid intelligence scores, and less likely to have a history of diabetes, hypertension, cerebrovascular diseases and depression at baseline (Supplementary Table 1). In addition, linear trends were also observed in the average consumption levels of the 34 food groups and component scores according to quartiles of MODERN diet score (Supplementary Table 9). In the fully adjusted model, participants in the highest quartile of the MODERN diet score had the lowest risk of ACD (HR for quantile 4 versus 1: 0.54; 95% CI: 0.45–0.65), with each 20% increment in the MODERN score associated with an 18% lower risk (HR : 0.82; 95% CI: 0.78–0.86). A similar trend was also found for AD (0.83; 95% CI: 0.76–0.89). The highest quartile of the MIND diet score was also associated, although less strongly, with a lower risk of ACD (HR_{Q4vsQ1} : 0.75; 95% CI: 0.65–0.87), with HR per 20% increment being 0.85 (95% CI: 0.79–0.91) (Extended Data Fig. 2 and Supplementary Table 10).

A series of sensitivity analyses demonstrated the robustness of the association of MODERN diet score with ACD and AD risk (Supplementary Table 11). The association remained largely similar after the exclusion of participants with onset in the first 2 or 5 years or additionally adjusting for a history of depression or fluid intelligence score at baseline and the intakes of the remaining 26 food groups (HR per 20% increment: 0.81–0.84). The modified MODERN diet including fish and other seafood, showed a similar association with ACD risk (HR per 20% increment: 0.82, 95% CI: 0.78–0.86), compared to the original pattern without this component. Incorporating fish and other seafood into the LightGBM model did not improve its discriminative performance (AUC: 0.719, 95% CI: 0.709–0.730) (Supplementary Table 7).

The inverse associations of MODERN diet score with ACD risk persisted in subgroups defined by age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity, BMI, *ApoE-ε4* (apolipoprotein E, type epsilon4 allele) gene, and history of major diseases except in non-whites and current smokers (Supplementary Table 11). Furthermore, the MODERN diet score was associated with a lower risk of ACD regardless of genetic risk defined by a polygenic risk score (PRS) for AD (Supplementary Table 12).

Next, we externally validated the observed association of the MODERN diet with the risk of dementia-related outcomes in the US-based Health and Retirement Study (HRS, N = 6,758, 58.7% females), Framingham Heart Study Offspring Cohort (FOS, N = 3,020, 54.6% females) and the National Health and Nutrition Examination Survey (NHANES, N = 2,3974, 53.3% females), totalling 33,752 individuals (Supplementary Table 13). Details of these cohorts are provided in Methods. We applied the identical MODERN diet score cut-offs developed in the UKB to these three US cohorts (HRS, FOS and NHANES) and assessed its associations with dementia-related outcomes (all-cause dementia in HRS and FOS, and AD mortality in NHANES) using the Cox proportional hazard models. The pooled HR for incident dementia comparing the highest versus the lowest tertile of the MODERN diet was 0.64 (95% CI: 0.43–0.93), which was stronger than that of the MIND diet (0.75, 0.61–0.92). Similarly, the HR per 20% increment was 0.74 (0.64–0.86) for MODERN and 0.78 (0.71–0.86) for MIND (Fig. 3c and Supplementary Table 14).

Outcome-wide association study of the MODERN diet score

To evaluate the potential effect of the MODERN diet score on other health outcomes, we utilized the first occurrence data in UKB and found that it was significantly associated with 49 out of 57 diseases after FDR correction (HR per 20% increment: 0.81–0.98; P_{FDR} : <0.001–0.036). Among these diseases, MODERN showed the strongest associations with mental and behavioural disorders (HR per 20% increment: 0.89; 95% CI: 0.87–0.90), especially psychotic disorders (HR per 20% increment: 0.81; 95% CI: 0.71–0.91). Although the association with all cancers was significant,

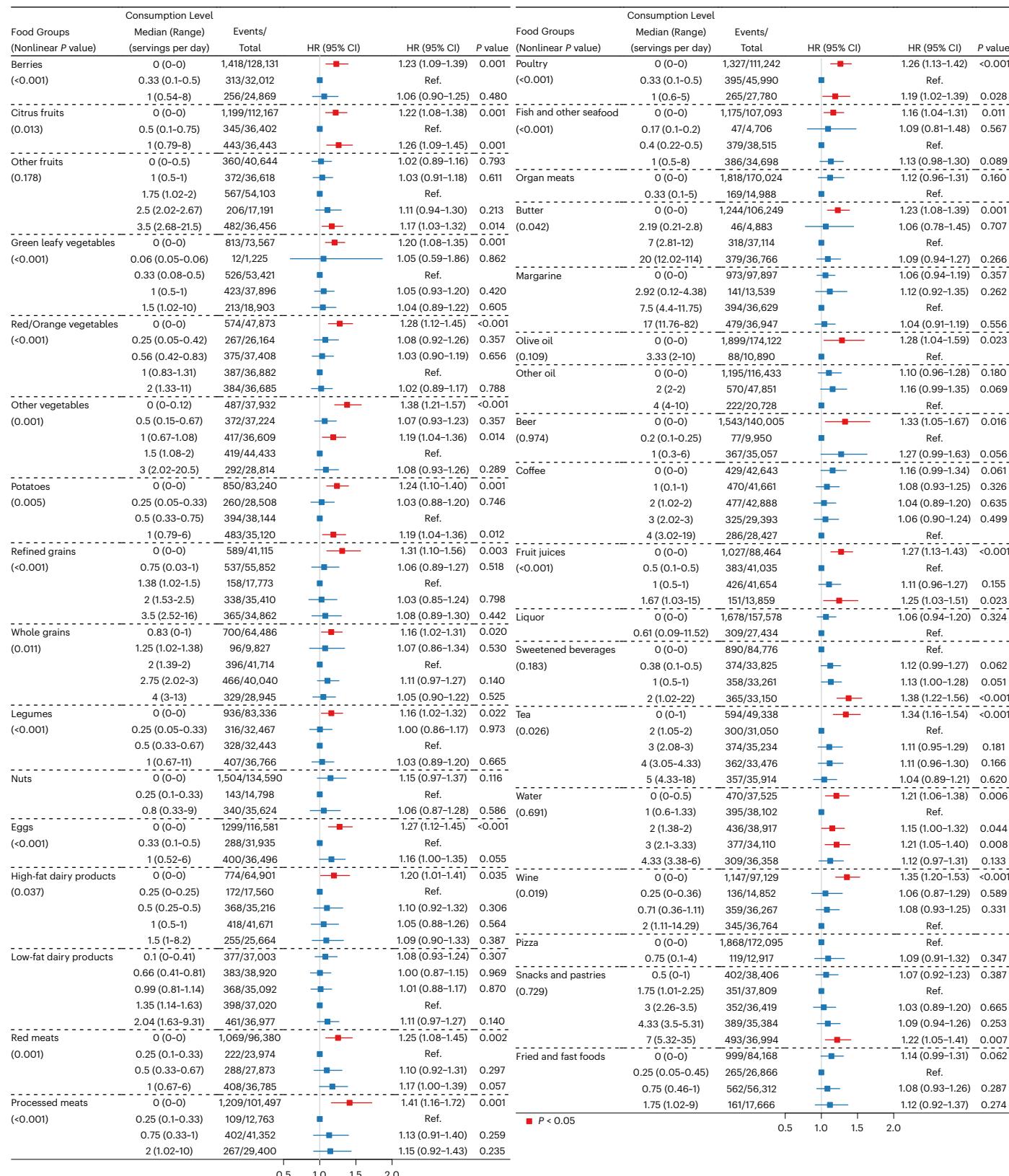


Fig. 2 | The associations of 34 food groups with incident ACD. Each food group was first grouped according to quintiles of consumption (or combined into one group if they were of the same zero quintile) and then included in the Cox proportional hazard regression models with different groups as references to make 2×2 comparisons. The figure only displays results of the groups with the lowest HR as a reference. A two-sided Wald test was performed to assess statistical significance. Significant associations ($P < 0.05$) are denoted in red in

the forest plot. A restricted cubic spline was also introduced to the models with 3 knots at the 10th, 50th and 90th percentiles, and the nonlinear P values were calculated to evaluate the nonlinear associations. The analyses were adjusted for total energy intake, age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity, BMI, *ApoE-e4* gene, and history of hypertension, diabetes, cerebrovascular disease and other CVDs.

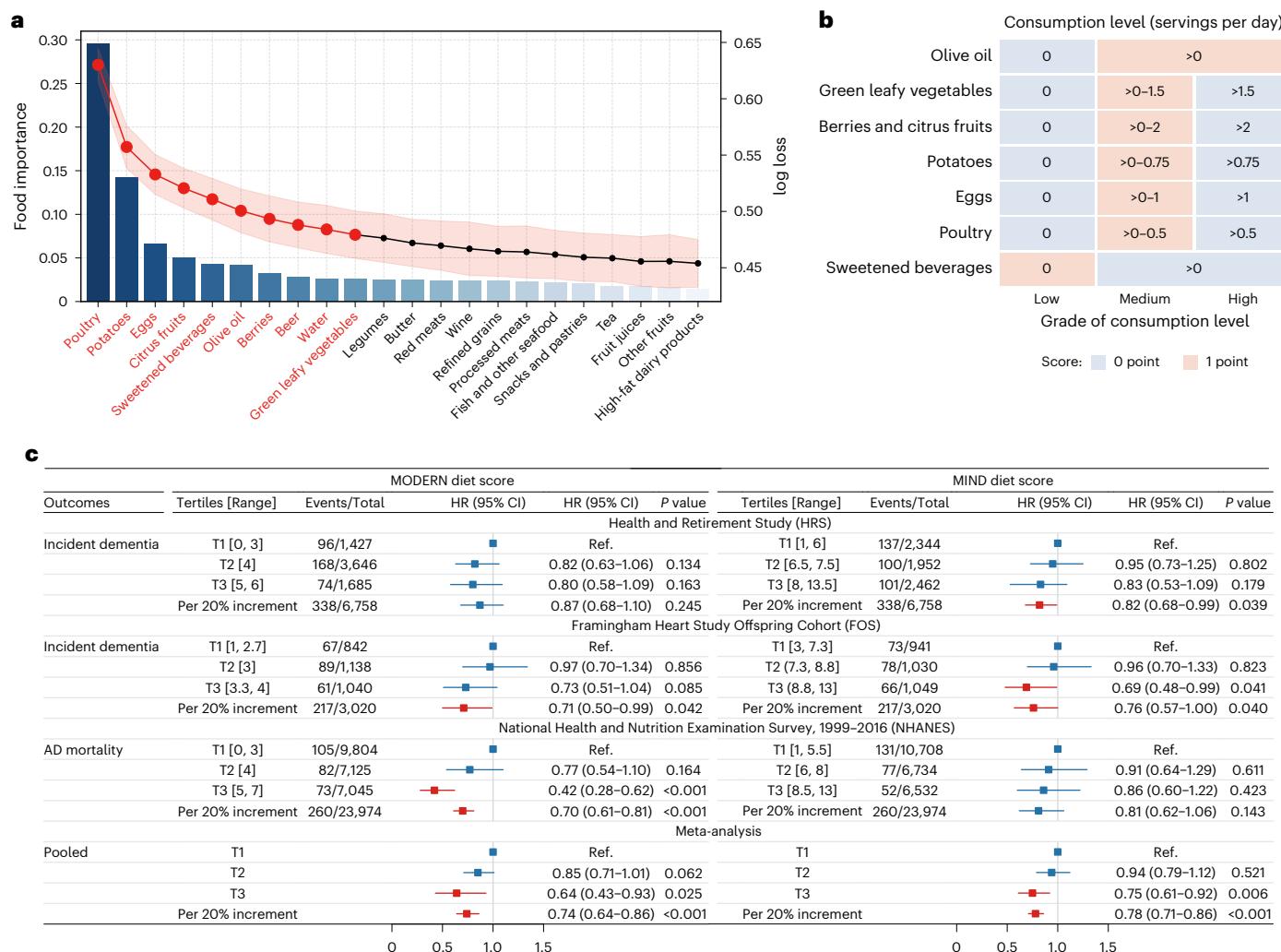


Fig. 3 | The MODERN diet score construction and its associations with dementia risk. **a**, Food group importance shown in the bar plot was generated from the LightGBM algorithm. The line chart shows the log loss as food groups were incrementally added to the model, with the shaded area representing the standard deviation. **b**, Eight food groups were selected to construct a new dietary pattern of seven components for dementia prevention, called MODERN diet. **c**, External validation in three independent external cohorts, showing the associations of MODERN and MIND diet with incident dementia/AD

mortality using the Cox proportional hazard regression model. Random-effects models were used to pool the risk estimates; $I^2 = 13.8\%$ and 0% for the estimates from MODERN and MIND diets, respectively. A two-sided Wald test was performed to assess statistical significance. Significant results ($P < 0.05$) are denoted in red. The analyses were adjusted for total energy intake, age, sex, ethnicity, income, education attainment, smoking status, physical activity, BMI, hypercholesterolaemia, and history of diabetes, hypertension, stroke, CVDs and depressive symptoms.

it was not in relation to most single cancer types except colorectal cancer, lung cancer and leukaemia. In addition, the MODERN diet score was associated with a lower risk of all-cause mortality ($HR_{per\ 20\% increment} = 0.91$, 95% CI: 0.89–0.93) and mortality due to cancer, nervous system diseases, circulatory system diseases and respiratory system diseases ($HR_{per\ 20\% increment} = 0.86-0.94$; $P_{FDR} < 0.001-0.010$; Fig. 4 and Supplementary Table 15).

The MODERN diet score and brain structure

Using brain magnetic resonance imaging data from a subset of UKB participants ($N = 21,584$), we observed significant associations between the MODERN diet score and the mean thickness of 31 out of 68 brain cortices, as well as the fractional anisotropy (FA) of 20 out of 27 white matter tracts and the mean diffusivity (MD) of 17 out of 27 white matter tracts after FDR correction (Fig. 5a,b and Supplementary Table 16). A higher MODERN score was associated with larger mean thickness of the superior frontal gyrus, middle frontal gyrus, inferior frontal gyrus, orbitofrontal cortex, paracentral lobule, precentral gyrus, postcentral gyrus, precuneus, superior temporal gyrus, transverse temporal gyrus,

middle temporal gyrus, entorhinal cortex, parahippocampal gyrus and insula (standardized coefficient $\beta: 0.013-0.035$; $P_{FDR}: <0.001-0.047$); and higher FA and lower MD of the thalamic radiation, corticospinal tract, inferior frontal-occipital tract, longitudinal fasciculus, uncinate fasciculus and forceps minor ($\beta: -0.028-0.031$; $P_{FDR}: 0.001-0.027$). In the analyses of food components (Fig. 5c), higher scores of berries and citrus fruits, and eggs were associated with most of the above significant thicknesses ($\beta: 0.017-0.028$; $P_{FDR}: <0.001-0.044$); and higher scores of green leafy vegetables and potatoes were associated with most of the above FAs ($\beta: 0.016-0.027$; $P_{FDR}: 0.001-0.025$), suggesting their potentially protective roles in maintaining brain structures.

Metabolomic, inflammatory and proteomic analyses

We observed significant associations between the MODERN diet score and 191 out of 249 circulating metabolites from 89,948 UKB participants ($\beta: -0.060-0.073$; $P_{FDR}: <0.001-0.045$; Fig. 5d and Supplementary Table 17) as well as 7 out of 9 inflammation markers from 169,115 UKB participants ($\beta: -0.006-0.024$; $P_{FDR}: <0.001-0.025$; Fig. 5e and Supplementary Table 17) after FDR correction. Notably, the most significant

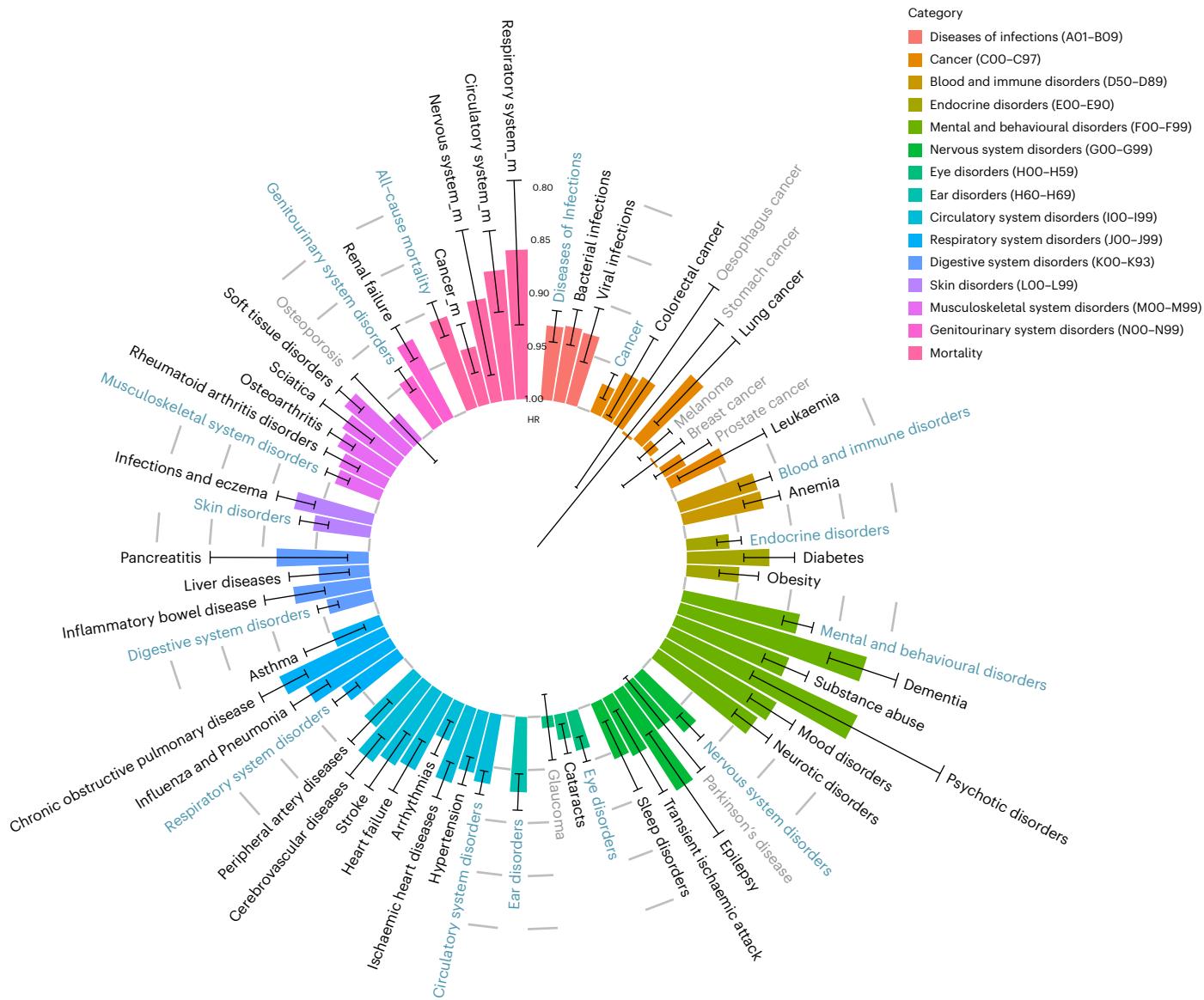


Fig. 4 | The associations of MODERN diet with other health-related outcomes. The analyses were conducted using the Cox proportional hazard regression model. Fifty-seven disease outcomes of 14 categories and 5 mortality outcomes were included according to the ICD-10 codes. The bar plot displays the HRs with each 20% increment in the MODERN score, and the error bars show the

95% confidence intervals. A two-sided Wald test was performed to assess statistical significance. Non-significant results after FDR correction are denoted in grey. The analyses were adjusted for total energy intake, age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity and BMI.

metabolites were n-3 polyunsaturated fatty acids (PUFAs), although fish and other seafood were not included in the dietary score. The MODERN-PUFA associations remained significant, although attenuated, after additionally adjusting for fish and other seafood, which suggests potential benefits of the MODERN diet independent of the intake of fish and other seafood. We conducted structural equation modelling (SEM) to evaluate the overall mediating effects of the top-10 most significant metabolites and the significant inflammation markers on the association between the MODERN diet score and ACD (Fig. 5f). We identified n-3 PUFAs to total fatty acid (TFA) percentage and docosahexaenoic acid (DHA) as the primary latent variables for metabolism. Systemic immune-inflammation index (SII) and neutrophil-to-lymphocyte ratio (NLR) were identified as the main latent variables for inflammation. The pathways of MODERN-metabolism-inflammation-ACD were significant (standardized coefficient: -0.043-0.126; $P < 0.001$ -0.004), suggesting that the MODERN diet is associated with a lower risk of ACD via metabolism-inflammation pathways.

Finally, we found that MODERN diet score was associated with 721 out of 2,919 plasma proteins in 16,899 UKB participants (β : -0.072-0.046; P_{FDR} : <0.001-0.049; Fig. 5g and Supplementary Table 18), among which 70 were also associated with the risk of ACD (β : -0.32-0.52; P_{FDR} : <0.001-0.049). In further mediation analyses, 40 proteins showed potential mediating effects, among which GFAP (8.4%), NEFL (6.2%), ELN (4.2%), TIMP4 (3.7%) and ADM (3.6%) showed potentially strongest effects (Fig. 5h and Supplementary Table 19). Gene ontology (GO) pathway enrichment analysis showed that these 40 proteins were enriched in neuron projection regeneration ($P_{\text{FDR}} < 0.001$). Protein-protein interaction analysis (PPI) revealed interactions between 26 proteins, among which EGFR had the most interactions with the remaining proteins, followed by TTR and GFAP (Fig. 5i).

Discussion

In the present study, we devised a new diet pattern for dementia prevention, named MODERN diet, by accounting for the longitudinal

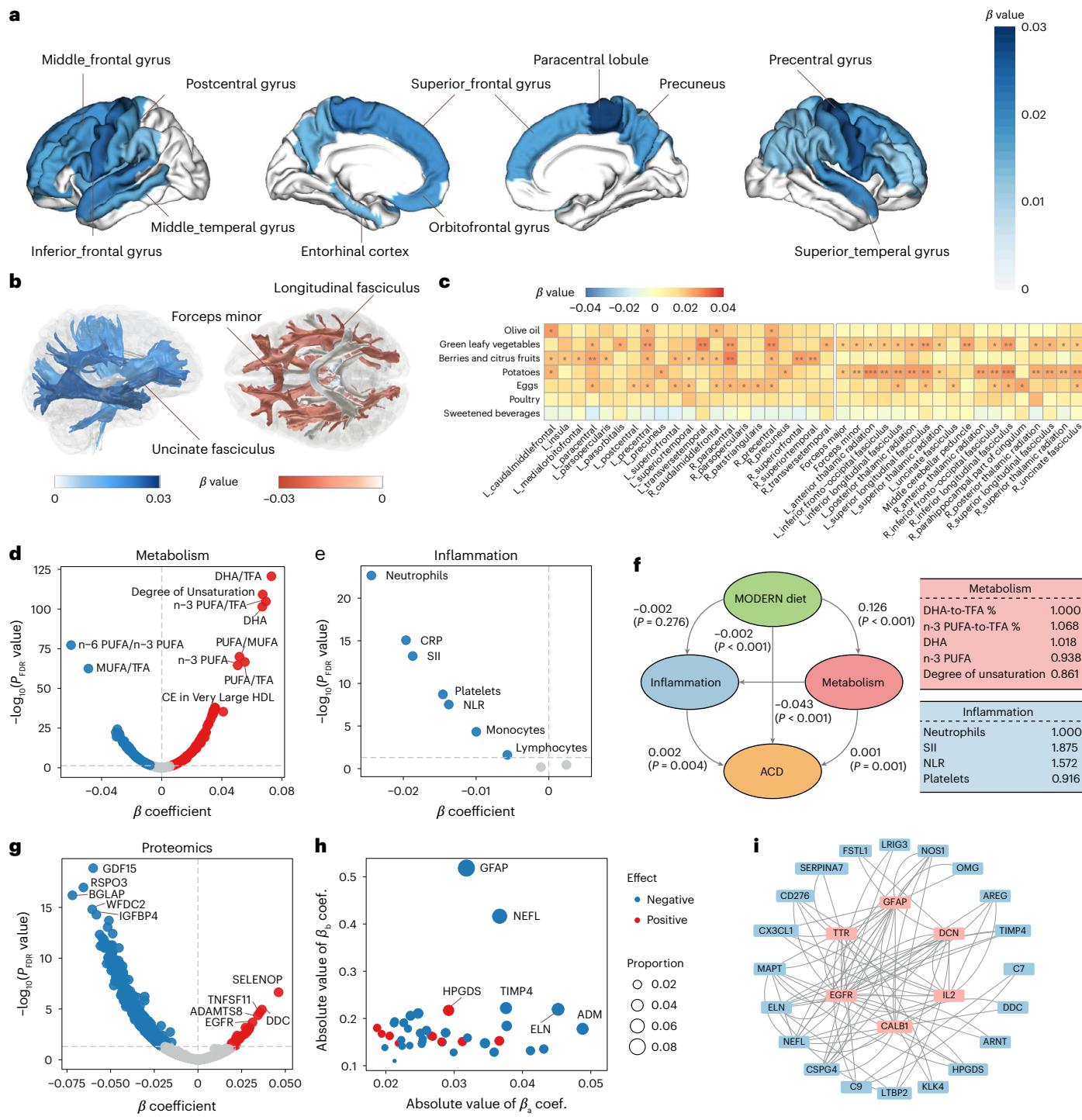


Fig. 5 | The potential pathways between the MODERN diet and ACD linking brain structure, metabolism, inflammation and proteomic signature.

a–c, Multiple linear regression models were used to estimate the associations of the MODERN diet with the thickness of 68 cortical regions (**a**), and the FA (left) and MD (right) of 27 white matter tracts (**b**). Similarly, the associations of the seven components with these metrics were estimated (**c**). FDR correction was used to account for multiple tests. * $P_{FDR} < 0.05$, ** $P_{FDR} < 0.01$ and *** $P_{FDR} < 0.001$. The β value is the standardized coefficient. **d,e**, Multiple linear regression models were used to estimate the associations of the MODERN diet with 249 metabolites

and 9 inflammation markers (**e**) in blood. **f**, SEM was used to estimate the mediating effects of metabolism and inflammation. The standardized coefficient and statistical significance are noted beside the path. **g**, Multiple linear regression models were used to estimate the associations between the MODERN diet and 2,919 plasma proteins. **h**, Mediation analyses were performed to find notable plasma proteins mediating the effect of the score on ACD incidence. Proteins with inverse associations with the score are denoted in blue. The size of the point indicates the proportion of the mediating effect. **i**, PPI analysis for the proteins with significant mediating effects.

and potentially nonlinear associations of various food groups with dementia risk. With an interpretable machine-learning approach, the MODERN diet emphasized the importance of 7 food components, including one adequacy component (>0 servings per day of olive oil),

five moderation components (0–1.5 servings per day of green leafy vegetables, 0–2 servings per day of berries and citrus fruits, 0–0.75 serving per day of potatoes, 0–1 serving per day of eggs, 0–0.5 serving per day of poultry) and one restriction component (0 serving per day

of sweetened beverages). It demonstrated a potentially stronger association than the previously established MIND diet (36% vs 25% lower risk of dementia comparing extreme tertiles) in the external validation. It also showed wide-range protective associations with other health outcomes, especially with mental and behavioural disorders. The connections of the MODERN diet score with larger mean thickness of the frontal, parietal and temporal cortices, higher FA and lower MD of multiple white matter tracts, lower level of inflammation, higher circulating n-3 PUFAs, and several dementia-related plasma proteins further support its potential benefits on brain health.

By incorporating an interpretable machine-learning approach based on a food-wide association analysis, the MODERN diet revised and added to the recommendations of the MIND diet to identify an optimized diet combination. MODERN diet recommended a higher intake of olive oil, as in the MIND diet. While green leafy vegetables and poultry were recommended for higher intake in the MIND diet, the MODERN diet recommended moderate intake. The MODERN diet also recommended moderate intake of fruits, and added citrus fruits to berries in fruit choices, both of which are rich in plant polyphenols and may improve cognitive function by inhibiting oxidative stress in the brain and neurodegenerative pathologies^{15–18}. Nevertheless, extremely high intakes of these components potentially increased the risk of incident dementia, in line with a previous UKB study suggesting that a ‘balanced’ diet derived from hierarchical clustering was related to better cognitive functions¹⁹. Although some longitudinal studies have reported modest associations between low-to-moderate alcohol consumption and cognitive outcomes²⁰, beer was not included due to potential confounding from health conditions and lifestyle factors, aligning with current World Health Organization (WHO) guidelines against any alcohol use²¹. Water was also excluded due to a lack of a well-established direct mechanistic link to dementia prevention in the literature. Notably, the MODERN diet does not include fish and other seafood, as their addition did not change the primary results or improve model discrimination performance when considered alongside other food groups. This may reflect inconsistencies in previous research findings and the context-specific nature of fish consumption—often fried in the United Kingdom—which warrants further investigation before being integrated into a brain-health-specific dietary pattern²².

Compared with the MIND diet, the MODERN diet additionally emphasized moderate intake of potatoes and eggs and limited intake of sweetened beverages on the basis of the findings from a food-wide association study that resonated with previous studies. In this study, potato, a starchy vegetable²³ rich in potassium, water-soluble vitamins and dietary fibre^{24,25}, is associated with lower dementia risk at a moderate intake level compared with zero intake. As for eggs, which showed inconsistent associations to various health outcomes, our study identified its potentially nonlinear association with dementia, similar to a Finland study²⁶. This is potentially because of its richness in choline and lutein^{27,28}. Notably, the MODERN diet restricted sweetened beverage (including sugar-sweetened or artificially sweetened beverages) intake, which was also supported by recent studies^{29–31} but not included in the MIND diet.

Consequently, we observed that the newly developed MODERN diet showed a consistent association with risk of dementia-related outcomes in three external cohorts, especially in the NHANES that employed a similar 24-h diet recall approach in assessing diet. Compared with participants in the lowest tertile of the MODERN diet score, those in the highest tertile exhibited a pooled 36% lower risk of dementia-related outcomes, which is stronger than that of the MIND diet (25% lower risk). Interestingly, we found the strongest associations of the MODERN diet with mental and behavioural disorders in addition to nervous system disorders in UKB, which suggested its higher relevance to neuropsychiatric health. This 7-item dietary pattern, if further validated among other population-based observational studies and RCTs, might be recommended to improve overall brain health

in older populations. With its wide range of health benefits beyond neurocognitive health, promoting this diet as part of public health initiatives could also be a cost-effective strategy for improving population health outcomes, although its efficacy and effectiveness both need to be examined further.

The potential cognitive benefits of the MODERN diet were further supported by the findings from multimodal analysis using neuroimaging, metabolomics, inflammation markers and plasma proteomics data. Dementia is characterized by thinning of the cortical thickness in the temporal, frontal and parietal lobe^{32,33}. Similar to the Mediterranean and MIND diets^{7,8}, MODERN showed associations with these brain structures, especially the middle temporal gyrus and entorhinal cortex, which have been described as the AD-signature regions⁸. Compared with our previous study based on hierarchical clustering of food preference data in UKB¹⁹, the MODERN diet was positively related to FA and negatively related to MD of more white matter tracts, such as the uncinate fasciculus and forceps minor, whose lesions are associated with cognitive dysfunction and neurodegenerative diseases^{34,35}. Inflammation is considered an important pathway between diet and dementia^{36,37}. In the present study, the MODERN–metabolism–inflammation–ACD pathway was prioritized by SEM. Higher circulating levels of n-3 PUFAs, especially DHA, induced by the MODERN diet, may maintain neuronal membrane functions^{38,39}. In addition, the MODERN diet may also suppress inflammation levels with its rich antioxidants such as green leafy vegetables, berries and citrus fruits, and olive oil^{36,37,40}. Interestingly, GFAP appeared to be the most relevant protein involved in the association between the MODERN diet and ACD. As a marker of astrocyte proliferation accompanying Aβ deposition^{41,42}, GFAP decreased after anti-Aβ monoclonal antibody treatment⁴³ and is considered a candidate for the prediction and early diagnosis of dementia and AD^{44,45}, which further supports the potential of the brain-targeting MODERN diet.

To the best of our knowledge, the current study is presumably the first to combine a food-wide longitudinal analysis and a machine-learning approach to develop a dietary pattern targeting dementia prevention. The multimodal data from UKB enabled us to explore the biological pathways with respect to brain structure, metabolism, inflammation and proteomic signature. Several limitations should be accounted for when interpreting our findings. First, 24-h diet recall may not fully capture long-term dietary behaviour and can lead to misclassification of less frequently consumed foods, which limit the interpretation of the findings. In this regard, the specific cut-offs of the devised MODERN diet warrant reevaluation in other studies, although the results from the food frequency questionnaire (FFQ)-based external validation cohorts (FOS and HRS) were similar as in the primary analyses based on UKB. Second, the ascensions of dementia and other endpoints are based on linkages to health systems, and underdiagnosis and misclassifications could not be ruled out. Third, this observational study does not eliminate the possibility of reverse causality and residual confounding, and future RCTs are needed to establish the causal associations between the newly devised dietary pattern and cognitive outcomes before any official recommendations or modifications to established diets can be made. Fourth, while the MODERN diet score was associated with lower dementia risk in UK-based and US cohorts, further research is needed to assess its applicability across a wider range of populations. Assessing its feasibility in international cohorts, particularly those from long-lived populations or blue zones, could provide additional insights into the generalizability of the dietary pattern. Future studies should aim to refine and validate the MODERN diet across diverse demographic and cultural contexts to ensure its global relevance in dementia prevention strategies. Finally, the investigation of the potential mechanism by multimodal analysis was based on the discovery cohort of UKB, and future studies are needed to verify the findings in external populations.

In conclusion, the current study devised a new dietary pattern for dementia prevention that showed a stronger association than a

previously established dietary pattern and may serve as a practical approach for the primary prevention of dementia. Future studies are warranted to validate this dietary pattern in diverse populations and evaluate its feasibility and effectiveness in clinical and public health practice.

Methods

Participants

The UK Biobank is a prospective population-based cohort study that recruited over half a million participants aged 37–73 years at baseline. At 22 assessment centres, information on a participant's demographic, lifestyle, health and physical assessments and questionnaires were collected from 2006 to 2010⁴⁶. One 24-h diet recall questionnaire was administered at the baseline assessment visit, followed by four online questionnaires sent via email during the 2011–2012 follow-up. From 210,926 participants who completed the questionnaire on at least one occasion, we invalidated questionnaires with completion period less than the 5th percentile (Field ID 20082) and excluded 23,782 participants who self-reported weekly changes in daily diet (Field ID 1548) and 2,067 participants who had doubtful estimates of total energy intake (<500 or >3,500 kcal per day for women and <800 or >4,200 kcal per day for men, Field ID 26002). In addition, 65 participants with dementia at baseline were excluded. The final sample size for the main analysis was 185,012 UKB participants. UKB was approved by the North West Multi-Center Research Ethics Committee (reference number 11/NW/03820). This study utilized the UKB resource under application number 19542. All participants provided written informed consent, and participants who dropped out were not included in this study. No direct financial compensation was provided to participants for their involvement.

Diet assessment

Diets were assessed using the Oxford WebQ questionnaire, which captures consumption of up to 206 foods and 32 beverages over the past 24 h^{47,48}, including a wide variety of vegetables, fruits, meats, dairy products, beverages and so on (UKB category 100090). The questionnaire asks about the consumption of a specific food in prespecified portion sizes, such as 'how many apples (fresh, frozen or canned) were eaten' with five responses ranging from 'half' to 'at least four'. Detailed questionnaire information is available at <https://www.ceu.ox.ac.uk/research/oxford-webq>. The portion size (1 serving) of each food item is specified in ref. 49 on the basis of how each question was asked and on the product information on packaging in different UK supermarkets. Previous studies investigating the reliability and internal consistency of the Oxford WebQ have shown that nutrient intakes calculated from this questionnaire correlate well with nutrient biomarkers measured from 24-h urine samples⁵⁰. For participants who completed more than one recall, the averages of each food were calculated, regarding the last time as the start of follow-up. In addition, the food frequency questionnaire administered at baseline was used to assess the reliability of the 24-h recalls (category 100052; Supplementary Table 20).

Outcomes

The primary outcome of the study was all-cause dementia (ACD), and the secondary outcome was Alzheimer's disease (AD). Dementia events were obtained from the first occurrence of health outcomes defined by 3-character International Classification of Diseases (ICD) codes (category 1712), which were generated from a wide range of health outcomes across self-report, primary care, hospital inpatient data and death data. Primary care data were obtained from various data suppliers and other intermediaries (including the main primary care computer system suppliers in England). Hospital inpatient data were collected from the Health Episode Statistics (HES) database for England and Wales and the Scottish Morbidity Records (SMR01) for Scotland. Information on the date and cause of death was recorded on

the National Health Service (NHS) Information Center and Central Register Scotland, respectively. More information can be found at <http://content.digital.nhs.uk/services>. Using the 10th version of ICD (ICD-10), ACD was defined as F00 (dementia in Alzheimer's disease), F01 (vascular dementia), F02 (dementia in other diseases) or F03 (unspecified dementia), and G30 (Alzheimer's disease), among which AD was F00 or G30. The end of the follow-up was the first occurrence of dementia, death, or the end of the study (31 December 2022).

Similarly, other health-related outcomes defined by ICD-10 were collected (category 1712), including 57 diseases across 14 categories depicted in previous studies⁵¹, as well as all-cause mortality and attributed mortality for cancer, neurologic diseases, circulatory diseases and respiratory diseases (Supplementary Table 21).

Covariates

At recruitment (2006–2010), participants provided data on demographic characteristics, including age, sex, educational attainment, area-based Townsend deprivation index and ethnicity, as well as lifestyle and genotype. Educational attainment was captured by the estimated number of years of education corresponding to self-reported educational qualifications⁵², which was then recoded into an ordinal variable from 0 to 10 (for example 0: 0 years, 10: >24 years)⁵³. The TDI is a measure of material deprivation, with lower scores reflecting higher affluence⁵⁴. Smoking status (never, previous or current) was derived from a touchscreen lifestyle questionnaire. According to a modified version of the International Physical Activity Questionnaire (IPAQ), physical activity was collected and transformed into a continuous variable (MET-minutes per week)⁵⁵. Body mass index was calculated by dividing weight (kg) by the square of height (m). *ApoE-ε4* genotype after quality control and imputation by UKB was coded according to the number of carried alleles (zero, one or two alleles). Further information on the genotyping process is available on the UKB website (<https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/genic-data>). Baseline health conditions, including hypertension (ICD-10: I10–13, I15), diabetes (E11–14), cerebrovascular diseases (I60–68); other cardiovascular diseases (CVDs) including coronary heart disease (I20–24), arrhythmia (I44–49), heart failure (I50) and peripheral vascular disease (I70–74, I77–82); and depression (F32–33) were obtained from the UKB first occurrence data (category 1712). From cognitive function tests administered at baseline, fluid intelligence score was employed to reflect baseline cognitive function (category 100026).

Optimizing dietary pattern determination

The optimized combination of a dietary pattern was determined in the derivation cohort of UKB in several steps. First, a food-wide association analysis was conducted to identify the food groups linearly or nonlinearly associated with the risk of ACD. Then, highly correlated food groups determined on the basis of Spearman correlations were removed after hierarchical clustering, and 22 food groups were left for further modelling. Next, a machine-learning approach was employed to prioritize food groups by trading off the quantity of included food groups versus their effectiveness in preventing dementia. To eliminate potential confounding bias by the risk factors of dementia, we standardized food group intake as residuals obtained from a linear regression accounting for total energy intake, age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity, BMI, *ApoE-ε4* gene, and history of hypertension, diabetes, cerebrovascular disease and other CVDs. We incorporated LightGBM⁵⁶ to determine the relative importance of the food groups quantified as information gain, derived from the entropy of tree splits during modelling, and it captured the extent to which each food group can impact the classification of dementia. Next, a sequential forward selection strategy was implemented by consecutively adding food groups one by one during each iteration and monitoring the decreasing trend of log loss within inner 10-fold cross-validation. The selection scheme was stopped when the

decrease in log loss in the consecutive food groups was less than 0.005. To enhance the interpretability of the black-box machine-learning models, we conducted SHapley Additive exPlanations (SHAP) analysis on the three models (LightGBM, XGBoost and Random Forest). Specifically, we averaged the absolute SHAP values from 10-fold inner cross-validation using the UKB data, and then selected the top-10 food groups on the basis of descending order of the derived values. To further validate the food groups selected in our approach, we also employed multiple machine-learning approaches; for example, importance ranking using entropy-based methods within XGBoost and random forest, and wrapper-based selection method of recursive feature elimination (RFE) using classifiers of LightGBM, XGBoost, random forest, logistic regression and support vector machine (SVM). Notably, the RFE set 10 predictors to be selected for comparison to our LightGBM-based methods. All machine algorithms, including LightGBM (`lightgbm v.3.3.2`), XGBoost (`xgboost v.5.1`), random forest, logistic regression, SVM and RFE (`scikit-learn v.1.2.2`), were implemented under the Python (`v.3.9`) environment. In addition, to evaluate and compare the discriminative capacity across different machine-learning models, we performed predictive analyses using multiple machine-learning algorithms based on different sets of selected food groups, namely, all 22 food groups, model-based top-10 food groups, SHAP value-based top-10 food groups, and RFE-based top-10 food groups. The analyses were performed using 10-fold cross-validation within the UK Biobank dataset, and we report predictive metrics of area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, accuracy, Youden index and area under the precision–recall curve.

Finally, we established a new diet score using the identified food groups, classified into three categories: adequacy, moderation and restriction. The adequacy component included food groups where higher intake was associated with lower dementia risk in a dose-response manner, and restriction components were defined reversely. Categories associated with the highest dementia risk were scored as 0, and all other levels received 1 point. The moderation component comprised those with a ‘U’-shaped relationship, where both low and high intakes were linked to higher risk. For these components, intake levels associated with the lowest and highest dementia hazards were assigned 1 and 0 points, respectively, while those that had a statistical difference from these reference points were scored as 0 or 1. The remaining intake levels, which were both significant or non-significant with the reference points, were further divided into two grades using the median as the threshold to be assigned to neighbouring levels, with the two ends assumed as 0. The scores for each component were then summed up as a composite diet score for dementia prevention.

The external validation cohorts

We included the Health and Retirement Study (HRS), the Framingham Heart Study Offspring Cohort (FOS), and the National Health and Nutrition Examination Survey (NHANES) in the United States for external validation.

The HRS is a longitudinal cohort study that surveys a representative sample of ~20,000 people in the United States⁵⁷. In 2013, the HRS Health Care and Nutrition Study (HCNS) utilized the Harvard FFQ to collect information about food and supplement intake, with a response rate of 65%. The HRS actively assessed the cognitive status of participants at biennial follow-ups, and we used the Langa–Weir algorithm to define prevalent and incident dementia cases⁵⁸.

The Framingham Heart Study is a community-based cohort study commenced in 1948, and the children of the original cohort and their spouses formed the FOS cohort in 1971. FOS participants completed dietary assessments in examinations 5 (1991–1995), 6 (1995–1998) and 7 (1998–2001) and have undergone continuous surveillance combined with active diagnosis of dementia through 2018.

Since 1999–2000, NHANES has enrolled non-overlapping participants every 2 years, collecting information on medical history, dietary

assessment, lifestyle and health conditions. Detailed information about NHANES has been described elsewhere⁵⁹. In NHANES, dietary intake was assessed using one (1999–2002) or the mean value of the two repeated (2003–2016) 24-h diet recalls, employing the US Department of Agriculture’s automated multiple-pass method, as described elsewhere⁶⁰. Given that incident dementia was not directly assessed in this cohort, the outcome of interest in the NHANES was AD mortality, which was ascertained with death certificate records by linkage to the National Death Index and represent a late stage of AD. The cause of death was coded according to ICD-10. The primary cause of death coded G30 was designated as AD-specific mortality (https://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/linked_mortality/).

For these external validation cohorts, we followed the inclusion/exclusion criteria described in our previous study². Participants younger than 45 years were excluded from the validation cohorts as they were less likely to develop dementia within the follow-up period. Thus, we included participants aged ≥45 years at baseline. In addition, we excluded those with unknown dementia status, baseline prevalent dementia, or incident dementia within the first 2 years of follow-up to minimize the potential for reverse causality. The flow chart illustrating participant inclusion is provided in Extended Data Fig. 3. We applied the identical MODERN diet score cut-offs developed in the UKB to these three US cohorts (HRS, FOS and NHANES) and assessed the associations using the Cox proportional hazard models. None of these cohorts were involved in the development or training of the MODERN score, ensuring independent validation. To account for the limited sample sizes within individual validation cohorts, we pooled estimates across the three cohorts with random-effect meta-analysis. We also assessed the associations for MIND diet for comparison. The study utilized the FOS data under application number 11068. All participants provided written informed consent. Financial compensation was provided to participants in the HRS and NHANES cohorts, but not for those in the FOS cohort.

MIND diet and scoring rules

The MIND dietary pattern was established in 2015 for neurodegenerative disease prevention on the basis of the Mediterranean diet and Dietary Approaches to Stop Hypertension (DASH) dietary pattern⁵. The MIND dietary pattern includes ten ‘brain healthy’ components and five ‘brain unhealthy’ components. The scoring rules were mainly based on previous literature⁴, with the recommended components, except for olive oil and wine, being scored 0 (low intake), 0.5 (medium intake) and 1 (high intake), while the non-recommended components were scored reversely. Olive oil was scored as 0 or 1 depending on whether it was used or not. Wine was scored as 0 when consumed at 0 or greater than 1 glass per day, 0.5 at 0–1 glass per day, and 1 at 1 glass per day (Supplementary Table 22). The composite score ranged from 0 to 15, and a higher score indicates higher adherence to the MIND diet.

Polygenic risk score

A weighted PRS was constructed using common genetic variants associated with AD identified in a previously published meta-analysis of genome-wide association study (GWAS) from 4 large AD consortia⁶¹. Because this score was based on a GWAS study of individuals of European ancestry, the present study was restricted to individuals whose self-reported ethnic background at baseline was white. Using a $P < 0.5$ as a threshold⁶², the most significant single-nucleotide polymorphisms (SNPs) in each block of linkage disequilibrium identified by the GWAS study were selected to establish a PRS for our population. Each SNP was weighted according to the strength of its association with AD and summed up to obtain the PRS for each individual. The PRS was then standardized and categorized into low (lowest quintile), intermediate (quintiles 2–4) and high (highest quintile) genetic risk. Plink 1.90 and PRSice (v.2.3.3) were used to generate the PRS.

Brain imaging

Since 2014, UKB has collected magnetic resonance imaging (MRI) data on the heads of over 40,000 UKB participants, all acquired by the same 3 T Siemens Skyra scanner with a 32-channel head coil. More details can be found at <https://www.fmrib.ox.ac.uk/ukbiobank/protocol/> (ref. 63). T1-weighted imaging (T1-MRI) was processed and analysed by the UKB imaging team using Freesurfer to generate different imaging-derived phenotypes (IDPs), including mean thickness, surface area, and volume of the cortical and subcortical regions (category 110). More details can be found at <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/> (ref. 64). Diffusion-weighted imaging (dMRI) techniques measure the ability of water molecules to move in localized tissue environments and can reflect microstructural integrity. After correcting for eddy currents and head movements, the dMRI data were used to generate microstructural phenotypes, such as the anisotropy (FA) and mean diffusivity (MD) of the white matter tracts, using the DTI Imaging Fitting Tool (DTFIT) (category 135). The full details of the image processing and quality control pipeline are available in an open-access Article⁶⁴. Ultimately, we included the mean thickness, surface area and volume of 68 cortical regions and the volume of 16 subcortical regions, as well as the FA and MD of 27 white matter tracts from 21,584 participants.

Inflammation markers and metabolites in plasma

At baseline 2006–2010, UKB collected blood specimens from ~500,000 people, more than 470,000 of whom had 31 routine blood items (category 100081) and 30 blood biochemistry items (category 17518) measured. Detailed steps for processing can be found in a UKB document at <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100080>. UKB performed a high-throughput nuclear magnetic resonance (NMR)-based metabolomic analysis of more than 270,000 randomly selected EDTA plasma from collected blood specimens (the analysis was performed by Nightingale Health, Helsinki, Finland), measuring a total of 249 metabolic metrics, including 168 biomarkers and 81 ratio parameters (category 220). These biomarkers cover multiple metabolic pathways, including lipoprotein lipids and various small molecules such as fatty acids, amino acids and glycolytic products. Metabolomic data from 89,948 participants were used in the present study. In addition, we incorporated neutrophils, monocytes, platelets and lymphocytes from routine blood tests and C-reactive protein (CRP) from blood biochemistry tests to estimate the level of inflammatory response in the body. On the basis of blood cell counts, we further calculated four ratios: neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), systemic immune-inflammation index (SII) and lymphocyte-to-monocyte ratio (LMR). Inflammation markers from 169,115 participants with 24-h diet recalls were used in the present study.

Olink plasma proteomic measurement

Proteomic biomarker data were produced using 55,000 samples collected from unique UKB participant visits, analysed over the period April 2021 to Feb 2022 using the Olink Explore platform, which covered up to 2,923 unique assays (category 1838)⁶⁵. Samples were processed across three NovaSeq 6000 Sequencing Systems. Extensive quality control measures and normalization of protein concentration were performed at Olink's facilities, producing normalized protein expression (NPX) values for each protein per participant. NPX is Olink's relative protein quantification unit on a log₂ scale⁶⁵. A summary of the quality control protocol can be found at http://biobank.ndph.ox.ac.uk/ukb/ukb/docs/PPP_Phase_1_QC_dataset_companion_doc.pdf. We excluded proteins with more than 25% of the samples below the lower limits of detection (LOD), resulting in a total of 2,919 proteins that were included in the analysis, covering four panels (cardiometabolic, inflammation, neurology and oncology). Proteomic data from 16,899 participants collected at baseline were used in the present study.

Statistical analyses

We collected 206 foods from the Oxford WebQ questionnaire that fell into the categories of fruits, vegetables, cereal and bread, legumes and nuts, dairy products and eggs, meats and seafood, fats and oils, water and beverages, snacks and pastries, and fast and fried foods. Canned or instant soups were not included due to the unavailability of specific ingredients in the foods. Also, dietary supplements such as vitamins were not included in the present study as its main focus was on people's daily diet. We categorized the 206 foods into 34 food groups mainly based on the methods described in refs. 14,66, as well as the food groups that are specifically related to brain health in previous literature and established dietary patterns such as MIND⁵. The consumption of each food group was summed up in 'servings'. For fats and oils, ref. 49 did not specify the size of a serving, and 10 g was regarded as a 'serving' according to British habits. Detailed information on these foods and food groups can be found in Supplementary Table 2. Their calculated average intakes were broadly in line with UK national intake levels (Supplementary Fig. 3). We analysed the Spearman correlations between them, which seemed generally weak ($R: -0.39\text{--}0.39$) except for that between vegetables ($R: 0.39\text{--}0.5$), and between coffee and tea ($R: -0.46$) (Supplementary Fig. 4). Dietary data measured by the FFQ at baseline and the 24-h diet recall questionnaire showed moderate correlations ($R: 0.23\text{--}0.81$) (Supplementary Fig. 5).

First, multivariable Cox proportional hazard regression models were employed to estimate the food-wide associations between the intake of various foods and food groups and the incidence of ACD. For different foods, we divided them into non-intake and intake groups, with the former as reference. For different food groups, they were divided into five levels according to quintiles (if the first few quintiles were the same zero, we combined them into one level). To better compare the different quintiles, we performed Cox regression analysis using the different levels as reference. Also, restricted cubic spline was introduced into the Cox regression model to assess possible nonlinear relationships, with 3 knots at the 10th, 50th and 90th percentiles. A similar analysis process was also conducted for the associations of the MODERN and MIND diet scores with incident dementia. The composite scores were divided into four or three groups on the basis of quartiles or tertiles in UKB cohorts and external cohorts, respectively, and the hazard ratio was calculated using the lowest quartile or tertile as reference. They were also included as a continuous variable in the Cox regression model to calculate the linear trend. The Cox proportional hazards models were evaluated for proportional hazard assumption violations using Schoenfeld residual tests ($P < 0.01$ divided by the number of multiple tests). We performed a two-sided Wald test to assess statistical significance. By incorporating different covariates, we created two regression models in the discovery cohort, the first model (Model 1, basic model) adjusting for total energy intake, age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity, BMI and *ApoE-ε4* gene; and the second (Model 2, full model) additionally adjusting for history of hypertension, diabetes, cerebrovascular disease and other CVDs. These covariates are either related to dietary intake or the incidence of dementia. The covariates were imputed using multiple imputations by chained equation, with 5 imputed datasets and 5 iterations⁶⁷. The covariates included in the validation cohorts almost covered these in Model 2, except for the *ApoE-ε4* gene as a previous study reported². Complete information regarding missing numbers is listed in Supplementary Table 23.

To verify the robustness of the results, we performed several sensitivity analyses. First, to minimize possible reverse causation, those with onset in the first 2 and 5 years were excluded. Second, to minimize the effects of baseline depressive status and cognitive function on the results, history of depression and fluid intelligence scores were used for further adjustment. Third, to attenuate the possible confounding effects of the remaining food groups, the other 26 food groups were used for further adjustment. Finally, we additionally included

fish and other seafoods as a component of the MODERN diet as it consistently ranked among the top-10 features in 6 of the 11 candidate machine-learning approaches. The predictive analysis was also conducted by incorporating fish and other seafood into the model-based top-10 food groups using the LightGBM algorithm. To evaluate the interaction of dementia-related environmental and genetic risk factors with the MODERN diet score, we conducted subgroup analyses stratified by age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity, BMI and *ApoE-ε4* gene, and major diseases, and statistically examined them by incorporating the interaction terms in the Cox regression model. In addition, genetic risk was further stratified according to the PRS for AD risk. Most importantly, to further validate the consistency of the score across different populations, external validation was conducted in three cohorts (HRS, FOS and NHANES) as mentioned above. To comprehensively evaluate the health benefits of the MODERN diet, we applied Cox regression analyses to calculate the effect of the score on other disease and mortality outcomes.

We used multiple linear regression models to calculate the associations of the MODERN diet score and each component score with the macro- and microstructure of the brain. The same analysis was employed for metabolites and inflammation markers in the blood. All independent and dependent variables were standardized to ensure comparability of coefficients. The statistical significance of individual coefficients was assessed using *t*-tests. Structural equation modelling was employed to determine the direction-dependent relationships mediated by inflammation and metabolism between the score and the ACD risk. Z-tests were used to assess the statistical significance of individual path coefficients. We also used multiple linear regression models to calculate the associations of the score with plasma proteins, and the logistic regression model was employed to calculate the association of the proteins with the risk of ACD. Finally, we identified the proteins that were significantly associated with both the score and the ACD risk. Mediation analysis was conducted to detect which among these proteins had mediating effects using quasi-Bayesian approximation. Subsequently, we used the STRING software to perform GO pathway enrichment analysis and protein–protein interaction analysis on the proteins with potential mediating effects⁶⁸. The covariates in Model 2 were used for adjustment in all association analyses, and scanning site was additionally used for adjustment in association analyses related to brain structures⁶⁹. For cortical thickness, area, volume and subcortical volume, the analysis additionally adjusted for total intracranial volume.

The statistical analyses were performed using R software (<http://www.r-project.org/>), with two-sided $P < 0.05$ as the threshold for statistical significance. False discovery rate (FDR) correction was used to account for multiple tests.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The main dataset supporting the conclusions of this article is available in the UK Biobank (UKB) repository (<https://www.ukbiobank.ac.uk/>). The disease and death outcomes in UKB can be obtained from the following restricted access national healthcare databases: the Hospital Episode Statistics (<https://digital.nhs.uk/services/hospital-episode-statistics>), the Scottish Morbidity Records (<https://www.ndc.scot.nhs.uk/National-Datasets/data-dictionary-smr01>), the National Health Service Information Center (<https://digital.nhs.uk>) and Central Register Scotland (<https://www.nrscotland.gov.uk>). This study utilized the UKB Resource under application number 19542. The Health and Retirement Study dataset is publicly available through its website (<https://hrsdata.isr.umich.edu/data-products/2013-health-care-and-nutrition-study-hcns>). The data from the Framingham Heart Study Offspring Cohort can be applied for at <http://www.framinghamheartstudy.org>.

[org/researchers/index.php](http://www.ncbi.nlm.nih.gov/researchers/index.php). This study utilized the FOS data under application number 11068. The National Health and Nutrition Examination Survey data are publicly available through the CDC/NCHS website (<https://www.cdc.gov/nchs/nhanes>). The summary statistics of the AD GWAS can be accessed at <https://gwas.mrcieu.ac.uk/datasets/ieu-b-2/>.

Code availability

The analysis programmes can be accessed on GitHub at https://github.com/Happychrischen/Diet_ML_Dementia (ref. 70).

References

1. Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446 (2020).
2. Chen, H. et al. Association of the Mediterranean Dietary Approaches to Stop Hypertension Intervention for Neurodegenerative Delay (MIND) diet with the risk of dementia. *JAMA Psychiatry* **80**, 630–638 (2023).
3. Zhang, Y. et al. Identifying modifiable factors and their joint effect on dementia risk in the UK Biobank. *Nat. Hum. Behav.* **7**, 1185–1195 (2023).
4. Morris, M. C. et al. MIND diet slows cognitive decline with aging. *Alzheimers Dement.* **11**, 1015–1022 (2015).
5. Morris, M. C. et al. MIND diet associated with reduced incidence of Alzheimer's disease. *Alzheimers Dement.* **11**, 1007–1014 (2015).
6. van den Brink, A. C., Brouwer-Brolsma, E. M., Berendsen, A. A. M. & van de Rest, O. The Mediterranean, Dietary Approaches to Stop Hypertension (DASH), and Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) diets are associated with less cognitive decline and a lower risk of Alzheimer's disease—a review. *Adv. Nutr.* **10**, 1040–1065 (2019).
7. Chen, H. et al. Associations of the Mediterranean-DASH Intervention for Neurodegenerative Delay diet with brain structural markers and their changes. *Alzheimers Dement.* **20**, 1190–1200 (2024).
8. Samuelsson, J. et al. Associations between dietary patterns and dementia-related neuroimaging markers. *Alzheimers Dement.* **19**, 4629–4640 (2023).
9. Agarwal, P. et al. Association of Mediterranean-DASH Intervention for Neurodegenerative Delay and Mediterranean diets with Alzheimer disease pathology. *Neurology* **100**, e2259–e2268 (2023).
10. Cornelis, M. C., Agarwal, P., Holland, T. M. & van Dam, R. M. MIND dietary pattern and its association with cognition and incident dementia in the UK Biobank. *Nutrients* **15**, 32 (2022).
11. Vu, T. H. T. et al. Adherence to MIND diet, genetic susceptibility, and incident dementia in three US cohorts. *Nutrients* **14**, 2759 (2022).
12. de Crom, T. O. E., Mooldijk, S. S., Ikram, M. K., Ikram, M. A. & Voortman, T. MIND diet and the risk of dementia: a population-based study. *Alzheimers Res. Ther.* **14**, 8 (2022).
13. Barnes, L. L. et al. Trial of the MIND diet for prevention of cognitive decline in older persons. *N. Engl. J. Med.* **389**, 602–611 (2023).
14. Piernas, C. et al. Describing a new food group classification system for UK Biobank: analysis of food groups and sources of macro- and micronutrients in 208,200 participants. *Eur. J. Nutr.* **60**, 2879–2890 (2021).
15. Zhang, S., Tomata, Y., Sugiyama, K., Sugawara, Y. & Tsuji, I. Citrus consumption and incident dementia in elderly Japanese: the Ohsaki Cohort 2006 Study. *Br. J. Nutr.* **117**, 1174–1180 (2017).
16. Zafra-Stone, S. et al. Berry anthocyanins as novel antioxidants in human health and disease prevention. *Mol. Nutr. Food Res.* **51**, 675–683 (2007).
17. Nakajima, A. et al. Nobiletin, a citrus flavonoid, ameliorates cognitive impairment, oxidative burden, and hyperphosphorylation of tau in senescence-accelerated mouse. *Behav. Brain Res.* **250**, 351–360 (2013).

18. Muhammad, T., Ikram, M., Ullah, R., Rehman, S. & Kim, M. Hesperetin, a citrus flavonoid, attenuates LPS-induced neuroinflammation, apoptosis and memory impairments by modulating TLR4/NF- κ B signaling. *Nutrients* **11**, 648 (2019).
19. Zhang, R. et al. Associations of dietary patterns with brain health from behavioral, neuroimaging, biochemical and genetic analyses. *Nat. Mental Health* **2**, 535–552 (2024).
20. Jeon, K. H. et al. Changes in alcohol consumption and risk of dementia in a nationwide cohort in South Korea. *JAMA Netw. Open* **6**, e2254771 (2023).
21. Rao, R. T. Methodological difficulties of studying alcohol consumption and dementia. *BMJ* **362**, k3894 (2018).
22. Zhang, Y. et al. Intakes of fish and polyunsaturated fatty acids and mild-to-severe cognitive impairment risks: a dose-response meta-analysis of 21 cohort studies. *Am. J. Clin. Nutr.* **103**, 330–340 (2016).
23. Wan, Y. et al. Association between changes in carbohydrate intake and long term weight changes: prospective cohort study. *BMJ* **382**, e073939 (2023).
24. Ayoob, K. T. Carbohydrate confusion and dietary patterns: unintended public health consequences of ‘food swapping’. *Front. Nutr.* **10**, 1266308 (2023).
25. Drewnowski, A., Maillot, M. & Vieux, F. Multiple metrics of carbohydrate quality place starchy vegetables alongside non-starchy vegetables, legumes, and whole fruit. *Front. Nutr.* **9**, 867378 (2022).
26. Ylilauri, M. P. et al. Association of dietary cholesterol and egg intakes with the risk of incident dementia or Alzheimer disease: the Kuopio Ischaemic Heart Disease Risk Factor Study. *Am. J. Clin. Nutr.* **105**, 476–484 (2017).
27. Institute of Medicine (US) Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and its Panel on Folate, Other B Vitamins, and Choline. *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline* (National Academies Press, 1998).
28. Vishwanathan, R., Kuchan, M. J., Sen, S. & Johnson, E. J. Lutein and preterm infants with decreased concentrations of brain carotenoids. *J. Pediatr. Gastroenterol. Nutr.* **59**, 659–665 (2014).
29. Chen, Y. et al. Associations of sugar-sweetened, artificially sweetened, and naturally sweet juices with Alzheimer’s disease: a prospective cohort study. *GeroScience* **46**, 1229–1240 (2023).
30. Liu, H. et al. Meta-analysis of sugar-sweetened beverage intake and the risk of cognitive disorders. *J. Affect. Disord.* **313**, 177–185 (2022).
31. Chen, H. et al. Sugary beverages and genetic risk in relation to brain structure and incident dementia: a prospective cohort study. *Am. J. Clin. Nutr.* **117**, 672–680 (2023).
32. Scheltens, P. et al. Alzheimer’s disease. *Lancet* **397**, 1577–1590 (2021).
33. Schliebs, R. & Arendt, T. The cholinergic system in aging and neuronal degeneration. *Behav. Brain Res.* **221**, 555–563 (2011).
34. O’Brien, J. T. Clinical significance of white matter changes. *Am. J. Geriatr. Psychiatry* **22**, 133–137 (2014).
35. Roseborough, A., Ramirez, J., Black, S. E. & Edwards, J. D. Associations between amyloid β and white matter hyperintensities: a systematic review. *Alzheimers Dement.* **13**, 1154–1167 (2017).
36. Peng, M. et al. Dietary inflammatory index, genetic susceptibility and risk of incident dementia: a prospective cohort study from UK Biobank. *J. Neurol.* **271**, 1286–1296 (2024).
37. Chen, X., Maguire, B., Brodaty, H. & O’Leary, F. Dietary patterns and cognitive health in older adults: a systematic review. *J. Alzheimers Dis.* **67**, 583–619 (2019).
38. Dyall, S. C. Long-chain omega-3 fatty acids and the brain: a review of the independent and shared effects of EPA, DPA and DHA. *Front. Aging Neurosci.* **7**, 52 (2015).
39. Chen, H. et al. Circulating metabolomic profile of the MIND diet and its relation to cognition in middle-aged and older adults. *iMetaOmics* **2**, e61 (2025).
40. McGrattan, A. M. et al. Diet and inflammation in cognitive ageing and Alzheimer’s disease. *Curr. Nutr. Rep.* **8**, 53–65 (2019).
41. Pereira, J. B. et al. Plasma GFAP is an early marker of amyloid- β but not tau pathology in Alzheimer’s disease. *Brain* **144**, 3505–3516 (2021).
42. Benedet, A. L. et al. Differences between plasma and cerebrospinal fluid glial fibrillary acidic protein levels across the Alzheimer disease continuum. *JAMA Neurol.* **78**, 1471–1483 (2021).
43. Pontecorvo, M. J. et al. Association of donanemab treatment with exploratory plasma biomarkers in early symptomatic Alzheimer disease: a secondary analysis of the TRAILBLAZER-ALZ Randomized Clinical Trial. *JAMA Neurol.* **79**, 1250–1259 (2022).
44. Huber, H. et al. Biomarkers of Alzheimer’s disease and neurodegeneration in dried blood spots—A new collection method for remote settings. *Alzheimers Dement.* **20**, 2340–2352 (2024).
45. Guo, Y. et al. Plasma proteomic profiles predict future dementia in healthy adults. *Nat. Aging* **4**, 247–260 (2024).
46. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
47. Liu, B. et al. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. *Public Health Nutr.* **14**, 1998–2005 (2011).
48. Bradbury, K. E., Young, H. J., Guo, W. & Key, T. J. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. *J. Nutr. Sci.* **7**, e6 (2018).
49. Perez-Cornago, A. et al. Description of the updated nutrition calculation of the Oxford WebQ questionnaire and comparison with the previous version among 207,144 participants in UK Biobank. *Eur. J. Nutr.* **60**, 4019–4030 (2021).
50. Greenwood, D. C. et al. Validation of the Oxford WebQ online 24-hour dietary questionnaire using biomarkers. *Am. J. Epidemiol.* **188**, 1858–1867 (2019).
51. Liu, W. et al. Association of biological age with health outcomes and its modifiable factors. *Aging Cell* **22**, e13995 (2023).
52. Cheesman, R. et al. Familial influences on neuroticism and education in the UK Biobank. *Behav. Genet.* **50**, 84–93 (2020).
53. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
54. Bond, J., Townsend, P., Phillimore, P. & Beattie, A. Health deprivation: inequality and the North, Croom Helm, Beckenham, 1987. *J. Soc. Policy* **18**, 291–293 (1989).
55. Craig, C. L. et al. International physical activity questionnaire: 12-country reliability and validity. *Med. Sci. Sports Exerc.* **35**, 1381–1395 (2003).
56. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. In *Proc. 31st Conference on Neural Information Processing Systems* (eds Guyon, I. et al.) 3147–3155 (NIPS, 2017).
57. Sonnega, A. et al. Cohort Profile: the Health and Retirement Study (HRS). *Int. J. Epidemiol.* **43**, 576–585 (2014).
58. Crimmins, E. M., Kim, J. K., Langa, K. M. & Weir, D. R. Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. *J. Gerontol. B* **66**, i162–i171 (2011).
59. NHANES Survey Methods and Analytic Guidelines (CDC, NCHS, year); <https://www.cdc.gov/Nchs/Nhanes/AnalyticGuidelines.aspx>
60. Ahluwalia, N., Dwyer, J., Terry, A., Moshfegh, A. & Johnson, C. Update on NHANES dietary data: focus on collection, release, analytical considerations, and uses to inform public policy. *Adv. Nutr.* **7**, 121–134 (2016).

61. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
62. Lourida, I. et al. Association of lifestyle and genetic risk with incidence of dementia. *JAMA* **322**, 430–437 (2019).
63. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
64. Alfaro-Almagro, F. et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424 (2018).
65. Wik, L. et al. Proximity extension assay in combination with next-generation sequencing for high-throughput proteome-wide analysis. *Mol. Cell. Proteomics* **20**, 100168 (2021).
66. Hu, F. B. et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am. J. Clin. Nutr.* **69**, 243–249 (1999).
67. Raghunathan, T., Lepkowski, J., Hoewyk, J. V. & Solenberger, P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* **27**, 85–95 (2001).
68. Szklarczyk, D. et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
69. Alfaro-Almagro, F. et al. Confound modelling in UK Biobank brain imaging. *Neuroimage* **224**, 117002 (2021).
70. Chen, S. J. Diet-ML-Dementia. Zenodo <https://doi.org/10.5281/zenodo.15671234> (2025).

Acknowledgements

We thank all the participants and professionals contributing to the UK Biobank and the Health and Retirement Study. This study was supported by grants from the Science and Technology Innovation 2030 Major Projects (2022ZD0211600 to J.-T.Y.), the National Natural Science Foundation of China (82071201, 82271471 and 92249305 to J.-T.Y.; 82071997 and 82472055 to W.C.; 8210120183 to C.Y.; 82402381 and 82471940 to J.Y.), the Shanghai Municipal Science and Technology Major Project (2023SHZDZX02 to J.-T.Y., 2018SHZDZX01 to J.-F.F.), the National Nutrition Science Research Grant (CNS-NNSRG2021-61 to C.Y.), the Zhejiang University Global Partnership Fund, Shanghai Pujiang Talent Program (23PJD006 to J.Y.), a Research Start-up Fund of Huashan Hospital (2022QD002 to J.-T.Y.), the Program of Shanghai Academic Research Leader (23XD1420400 to J.-T.Y.), the Excellence 2025 Talent Cultivation Program at Fudan University (3030277001 to J.-T.Y.), Shanghai Talent Development Funding for The Project (2019074), Shanghai Rising-Star Program (21QA1408700 to W.C.), 111 Project (B18015 to J.-F.F.), the National Postdoctoral Program for Innovative Talents (BX20230087 to S.-D.C.), the China Postdoctoral Science Foundation (2023M740672 to S.-D.C.), and ZHANGJIANG LAB, Tianqiao and

Chrissy Chen Institute, the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of the Ministry of Education, and Shanghai Center for Brain Science and Brain-Inspired Technology, Fudan University. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

Author contributions

J.-T.Y. and C.Y. had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. J.-T.Y. conceptualized and designed the project. All authors acquired, analysed or interpreted data. S.-J.C., S.-D.C., H.C., J.Y., C.Y. and J.-T.Y. drafted the paper. S.-J.C., H.C., J.Y., S.-D.C., L.H., X.G., W.C., C.Y. and J.-T.Y. critically revised the paper for important intellectual content. S.-J.C., J.Y., H.C., Y.F., W.Z. and W.C. performed statistical analysis. J.Y., S.-D.C., J.-F.F., W.C., C.Y. and J.-T.Y. obtained funding. J.-F.F., W.C., C.Y. and J.-T.Y. provided administrative, technical or material support. All authors read and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41562-025-02255-w>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-025-02255-w>.

Correspondence and requests for materials should be addressed to Changzheng Yuan or Jin-Tai Yu.

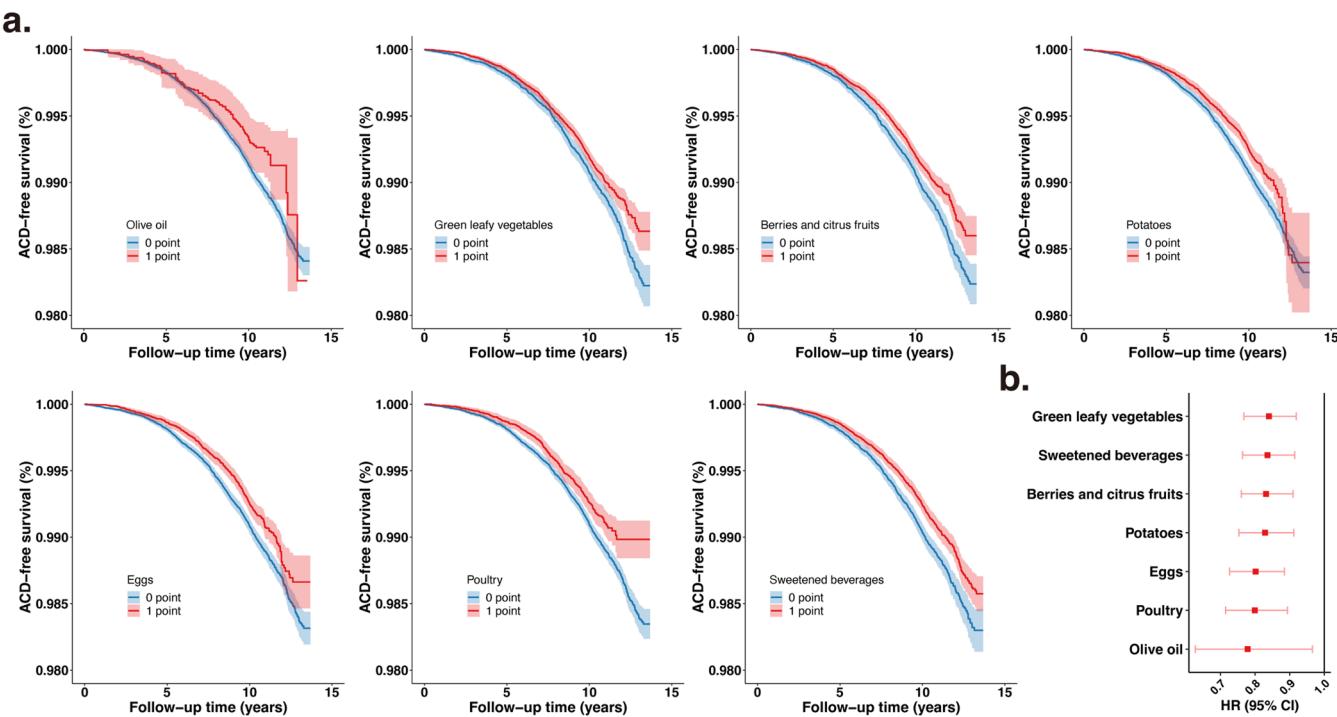
Peer review information *Nature Human Behaviour* thanks Rob van Dam and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

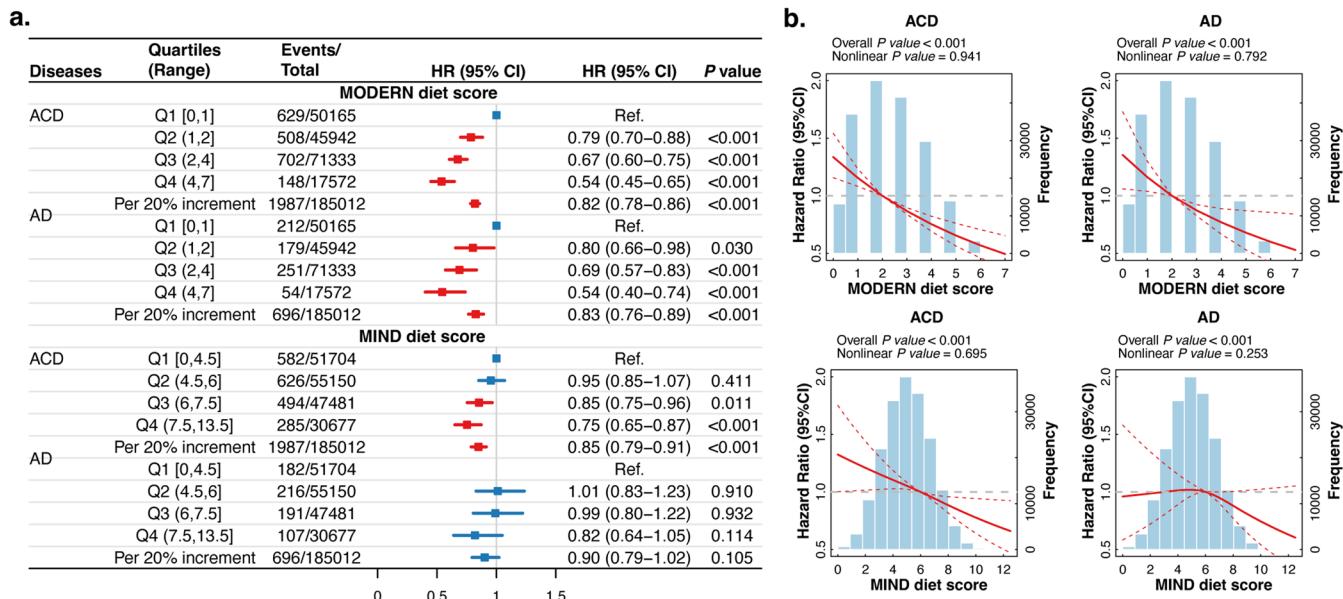
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025



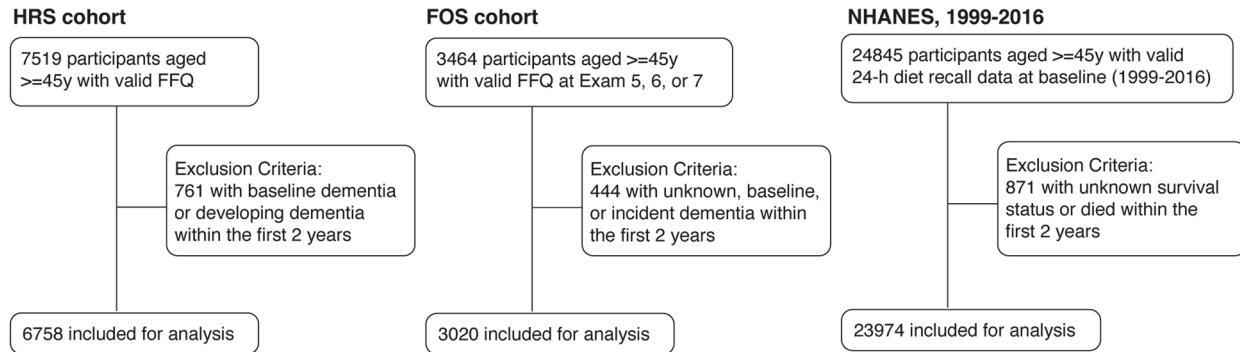
Extended Data Fig. 1 | The associations between the components of the MODERN diet and ACD risk in 185,012 UKB participants. **a.** The Kaplan-Meier curves of ACD-free survival in participants receiving 0 vs. 1 point for each food component. Shaded areas showed 95% confidence intervals. **b.** The HR for incident ACD comparing 1 vs. 0 point was estimated using the Cox proportional hazard regression model. A two-sided Wald test was performed to assess

statistical significance. The significant results (P value < 0.05) were denoted in red in the forest plot. The analyses were adjusted for total energy intake, age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity, BMI, *ApoE-ε4* gene, and history of hypertension, diabetes, cerebrovascular disease, and other CVDs. ACD: all-cause dementia; TDI: Townsend deprivation index; BMI: body mass index; CVDs: cardiovascular diseases.



Extended Data Fig. 2 | The associations between the MODERN diet and ACD risk in 185,012 UKB participants. **a.** The associations of MODERN and MIND diet with incident ACD using the Cox proportional hazard regression model. **b.** The nonlinear associations between MODERN and MIND diet score and incident dementia using the RCS. A two-sided Wald test was performed to assess statistical significance. The significant results (P value < 0.05) were denoted in red in the forest plot. The dashed lines showed 95% confidence intervals.

The analyses were adjusted for total energy intake, age, sex, ethnicity, TDI, educational attainment, smoking status, physical activity, BMI, *ApoE-ε4* gene, and history of hypertension, diabetes, cerebrovascular disease, and other CVDs. MIND: Mediterranean-DASH Intervention for Neurodegenerative Delay; ACD: all-cause dementia; RCS: restricted cubic spline; TDI: Townsend deprivation index; BMI: body mass index; CVDs: cardiovascular diseases.



Extended Data Fig. 3 | The inclusion and exclusion criteria of participants in external cohorts. HRS: Health and Retirement Study. FOS: Framingham Heart Study Offspring Cohort. NHANES: National Health and Nutrition Examination Survey.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was involved in data collection from UK Biobank, the Framingham Heart Study Offspring Cohort (FOS), and the National Health and Nutrition Examination Survey (NHANES). The Langa-Weir algorithm was adopted to define prevalent and incident dementia cases based on the cognitive status of participants in the Health and Retirement Study (HRS).
Data analysis	R version 4.3.1 packages: survival 3.5.5 was used to perform Cox proportional hazard regression model in the longitudinal analysis; mice 3.16.0 was used for data imputation; psych 2.3.9 was used to calculate the Spearman correlations between the intake of food groups; rms 6.7.1 was used to estimate the nonlinearity of the food-dementia associations; mediation 4.5.0 was used to perform mediation analysis; lavaan 0.6.16 was used to perform structural equation model. python (v3.9), lightgbm library (v3.3.2), xgboost library (v1.5.1), scikit-learn library (v1.2.2) was used for machine learning; Plink 1.90 and PRSice (v2.3.3) was used to for the generation of AD-PRS; STRING (https://www.stringdb.org/) was used for protein enrichment analysis and protein-protein interaction;

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The data used in the present study are available from UK Biobank (UKB) with restrictions applied. Data were used under license and are thus not publicly available. Access to the UKB data can be requested through a standard protocol (<https://www.ukbiobank.ac.uk/register-apply/>). The disease and death outcomes in UKB can be obtained from the following restricted access national healthcare databases: the Hospital Episode Statistics (<https://digital.nhs.uk/services/hospital-episode-statistics>), the Scottish Morbidity Records (<https://www.ndc.scot.nhs.uk/National-Datasets/data-dictionary-smr01/>), the National Health Service Information Center (<https://digital.nhs.uk>) and Central Register Scotland (<https://www.nrscotland.gov.uk>). The data from Health and Retirement Study can be accessed from its website (<https://hrsdata.isr.umich.edu/data-products/2013-health-care-and-nutrition-study-hcns>). The data from the Framingham Heart Study Offspring Cohort can be applied at <http://www.framinghamheartstudy.org/researchers/index.php>. The National Health and Nutrition Examination Survey data is publicly available through CDC/NCHS website (<https://www.cdc.gov/nchs/nhanes/>). The summary statistics of AD GWAS can be accessed from <https://gwas.mrcieu.ac.uk/datasets/ieu-b-2/>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

We took sex into considerations in our study and our findings could apply to both man and woman. Sex (Field ID 31) in the UK Biobank was determined based on self-reporting data via questionnaire, and all included participants gave written informed consent for sharing of individual-level data. The external validation cohorts also took sex into considerations.

Reporting on race, ethnicity, or other socially relevant groupings

We took ethnicity into considerations as one population characteristic and the discovery cohort was mainly based on white populations. Ethnicity (Field ID 21000) in the UK Biobank was determined based on self-reporting data via questionnaire, and all included participants gave written informed consent for sharing of individual-level data. The external validation cohorts also took ethnicity into considerations.

Population characteristics

This study included 185012 UK Biobank adults (102304 [55.3%] women; mean [SD] age, 59.2 [7.92] years). Over a mean of 10 years of follow-up, 1987 participants (1.07%) developed dementia, and the mean (SD) age at diagnoses were 74.6 (5.4) years. The baseline demographic data of participants is shown in Supplementary Table 1. We calculated descriptive statistics as mean (SD) for continuous variables and number (percentage) for categorical variables. The external validation included three US-based cohorts, namely the Health and Retirement Study (HRS, N=6758, 58.7% females), Framingham Heart Study Offspring Cohort (FOS, N=3020, 54.6% females), and the National Health and Nutrition Examination Survey (NHANES, N=23974, 53.3% females), totaling 33,752 individuals. The baseline demographic data of participants is shown in Supplementary Table 12.

Recruitment

The UK Biobank enrolled the participants aged 40-69 years between 2006 and 2010 for baseline assessments in 22 centers across the UK. The assessment visits comprised interviews and questionnaires covering lifestyles and health conditions, physical measures, biological samples, imaging, and genotyping. The 24-hour diet recall questionnaire was administered at the baseline assessment visit, followed by four online questionnaires at home during the 2011-2012. The database is linked to national health datasets, including primary care, hospital inpatient, death, and cancer registration data. The Health and Retirement Study (HRS) is a longitudinal cohort study that surveys a representative sample of approximately 20,000 people in the United States. A Havard food frequency questionnaire was administered in 2013 and the cognitive status of participants was evaluated at biennial follow-ups. The Framingham Heart Study is a community-based cohort study commenced in 1948, and the children of the original cohort and their spouses formed the Framingham Heart Study Offspring Cohort (FOS) cohort in 1971. FOS participants completed dietary assessments in examinations 5 (1991-1995), 6 (1995-1998), and 7 (1998-2001) and have undergone continuous surveillance combined with active diagnosis of dementia through 2018. National Health and Nutrition Examination Survey (NHANES) has enrolled non-overlapping participants every two years since 1999-2000, collecting information on medical history, dietary assessment, lifestyle, and health conditions. Dietary intake was assessed using 24-hour diet recalls employing the United States Department of Agriculture's (USDA's) Automated Multiple-Pass Method. The outcome of AD mortality was ascertained with death certificate records by linkage to the National Death Index.

Ethics oversight

The UK Biobank has received ethical approval from the North West Multi-centre Research Ethics Committee (MREC, <https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics>), and informed consent through electronic signature was obtained from study participants. This study utilized the UK Biobank Resource under application number 19542. The Health and Retirement Study was approved by University of Michigan institutional review board. The data from the Framingham Heart Study Offspring Cohort can be applied at <http://www.framinghamheartstudy.org/researchers/index.php>.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes. Out of 210926 UK Biobank (UKB) participants with 24-hour recall diet data, 185012 were eligible and were included for analysis after excluding 23782 participants who self-reported weekly changes in daily diet (Field ID 1548), 2,067 participants who had doubtful estimates of total energy intake (Field ID 26002), and 65 with baseline dementia. In external validation cohorts, participants younger than 45 years were excluded from the validation cohorts. We excluded those with unknown dementia status, baseline prevalent dementia, incident dementia, or died within the first two years of follow-up. Finally, a total of 33,752 individuals were included. Although no sample size calculation was conducted, our sample (in both UKB and external cohorts) was supported by precedent from similar published studies (DOI:10.1001/jamapsychiatry.2023.0800).
Data exclusions	In UK Biobank, participants who self-reported weekly changes in daily diet (Field ID 1548), had doubtful estimates of total energy intake (Field ID 26002), and had baseline dementia were excluded. In external validation cohorts, participants younger than 45 years or with unknown dementia status, baseline prevalent dementia, incident dementia or died within the first two years of follow-up were excluded.
Replication	All available UKB data were used to maximize statistical power of the analysis therefore we did not repeat the analysis. The data from Health and Retirement Study (HRS), the Framingham Heart Study Offspring Cohort (FOS) and the National Health and Nutrition Examination Survey (NHANES) were used for external validations.
Randomization	Covariates including total energy intake, age, sex, ethnicity, Townsend deprivation index (TDI)/income, educational attainment, smoking status, physical activity, BMI, ApoE-ε4 gene, and history of hypertension, diabetes, cerebrovascular disease, and other cardiovascular diseases.
Blinding	Blinding was not applicable to this study as this study is observational.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants		

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

Magnetic resonance imaging

Experimental design

Design type	Linear regression models using phenotypes derived from T1-MRI and Diffusion MRI
Design specifications	The UK Biobank designed the imaging acquisition protocols including 6 modalities, covering structural, diffusion and functional imaging. In the current study, T1-weighted structural image provides a high-resolution depiction of brain anatomy. Diffusion MRI (voxel-wise) was used to reflect the integrity of microstructural tissue compartments.
Behavioral performance measures	The associations between a novel dietary pattern score and brain structures derived from MRI were estimated. This dietary pattern consisted of 7 food components, including one adequacy components (olive oil), five moderation components (green leafy vegetables, citrus fruits and berries, potatoes, eggs, and poultry), and one restriction components (sweetened beverages).

Acquisition

Imaging type(s)	T1-weighted structural imaging
Field strength	3T
Sequence & imaging parameters	Resolution: 1x1x1 mm Field-of-view: 208x256x256 matrix Duration: 5 minutes 3D MPRAGE, sagittal, in-plane acceleration iPAT=2, prescan-normalise
Area of acquisition	The whole brain were acquired.
Diffusion MRI	<input checked="" type="checkbox"/> Used <input type="checkbox"/> Not used
Parameters	Resolution: 2x2x2 mm Field-of-view: 104x104x72 matrix Duration: 7 minutes (including 36 seconds phase-encoding reversed data) 5x b=0 (+3x b=0 blip-reversed), 50x b=1000 s/mm ² , 50x b=2000 s/mm ² Gradient timings: δ=21.4 ms, Δ=45.5 ms; Spoiler b-value = 3.3 s/mm ² SE-EPI with x3 multislice acceleration, no iPAT, fat saturation For the two diffusion-weighted shells, 50 distinct diffusion-encoding directions were acquired.

Preprocessing

Preprocessing software	T1-MRI was processed and analyzed by the UK Biobank imaging team using Freesurfer to generate macrostructural phenotypes. Diffusion MRI was used to generate microstructural phenotypes using the DTI Imaging Fitting Tool (DTFIT). The full details of the imaging processing and quality control pipeline are available in an open-access article 10.1016/j.neuroimage.2017.10.034.
Normalization	see above
Normalization template	see above
Noise and artifact removal	see above
Volume censoring	see above

Statistical modeling & inference

Model type and settings	Linear regression models.
Effect(s) tested	The regression coefficient was obtained.
Specify type of analysis:	<input checked="" type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference	Voxel-wise
(See Eklund et al. 2016)	
Correction	False discovery rate (FDR)

Models & analysis

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input checked="" type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis