SUPPLEMENTARY MATERIAL

eMethods

Neuroimaging protocol

MRI acquisition was performed on 1.5-T Magnetom (Siemens, Erlangen, Germany), on average 0.3 years after baseline (range, 0 to 1.3 years). A 3D high-resolution T1-weighted brain image was acquired using a 3D inversion-recovery fast-spoiled gradient-echo sequence (3D IRSPGR; repetition time [TR] = 97 ms; echo-time [TE] = 4 ms; inversion time [TI] = 600 ms; coronal acquisition). The axially reoriented 3D volume matrix size was 256×192×256 with a 1.0×0.98×0.98 mm^3 voxel size. There were 124 slices covering the whole brain. T2 and proton density (PD) weighted brain volumes were acquired using a 2D dual spin-echo sequence with two echo times (TR = 4400 ms; TE1 = 16 ms; TE2 = 98 ms). T2 and PD acquisitions consisted of 35 axial slices 3.5 mm thick (0.5 mm between slices spacing), having a 256×256 matrix size, and a 0.98×0.98 mm^2 in-plane resolution. Datasets (T1, T2 and PD) were readily reconstructed and visually checked for major artifacts.

Statistical analyses

Factorial analysis of mixed data

Factorial Analysis of Mixed Data (FAMD) is a dimension-reduction approach close to principal component analysis (PCA). The general goal of these approaches is to reduce the dimension of a dataset into a few summary variables, defined as linear combinations of initial predictors, which explain as much as possible of the variance of data. While PCA is adapted to continuous variables, FAMD allows a mixture of continuous and categorical variables. FAMD combines PCA (for continuous variables) with multiple component analysis used for categorical variables, and balances the influence of both types of variables in the dimension reduction procedure. In FAMD, continuous variables are scaled to a unit variance (as in PCA), while categorical variables are transformed into a disjunctive table (i.e. one column per modality) and each resulting modality is scaled to take into account the number of individuals presenting that modality (as in multiple component analysis). Principal components were identified by orthogonal rotation, which allows direct interpretation of the components loadings in terms of correlation between variables and dimensions.[1]

As with other reduction-dimension approaches, in FAMD, the linear combination of initial predictors that define each *principal component* is the sum of each predictor weighted by a specific *coefficient*. The coefficients associated with each predictor (continuous variables or modalities for categorical variables) to define each component are key for the interpretation of that component. Indeed, predictors that have high (absolute) loads have an important contribution to the component.

In reduction-dimension approaches, a *variable space* is defined by *p* variables and characterized by geometrically defined *eigenvectors* using singular value decomposition. A change of basis is then applied to the variable space, by sequentially identifying the eigenvectors (orthogonal to each other), to construct a *factorial space* (with a maximum of *p* dimensions) that best summarize the data while keeping most of its information. *Factorial*

planes can be defined by two dimensions of the factorial space; and variables/modalities or observations can be projected onto it.

Specifically for FAMD, the resulting *factorial coordinates* are computed from different methodologies for continuous and categorical variables, so that they are not directly comparable. For continuous variables, the coordinate is the linear correlation between the variable and the component, as in PCA; and projections of variables onto factorial planes can be visualized with correlation circles. For categorical variables, the modality's coordinate is the barycenter of projected observations which possess the modality normalized by the inverse of the modality frequency, as in multiple component analysis; and projections of modalities onto factorial planes can be visualized by factorial maps.

There is a direct link between factorial coordinates of each variable in the factorial space, eigenvectors from the variable space, and the coefficients of the linear combinations of variables for definition of principal components.

Eigenvector for each principal component is defined by the factorial coordinates of each variables/modalities divided by the squared root of the variance explained by the component (i.e., the component's eigenvalue)

In FAMD, the *coefficients* (or loadings) are computed to take into account the difference in methodologies used for coordinate calculation. For continuous variables, coefficients are calculated as the variable eigenvector element divided by the empirical standard deviation of the variable. For categorical variables, coefficients are calculated for each modality (coded as 0/1 for presence/absence) as the corresponding eigenvector element.

FAMD principal components are then defined as the linear combination of each variable/modality weighted by its corresponding coefficient.

The FAMD was conducted using "PCAmix" function from the R package "PCAmixdata".[2]

Supplementary analyses

We examined the robustness of our results to any selection bias (hat may arise because participants included in the analysis are different from those not included) using a propensity score weighted analysis,[3,4] where individuals are weighted by a propensity score assessing the probability of being included in the analysis.

The propensity score of each participant was obtained by fitted values of a multivariable logistic regression modeling inclusion status (yes/no) as a dependent variable and including as independent variables: study exposure (fish intake); covariates used in the primary analysis; as well as a z-score for global cognition (combining the five cognitive tests: Mini-Mental State Evaluation, Benton Visual Retention Test, Isaacs Set Test, Trail Making Test A and B) which was associated with the inclusion status (**Table e-1**).[4] The adjusted linear regression between fish intake and the indicator of cerebrovascular burden was then weighted by the individuals' inverse of the propensity score. As in the main analysis, missing data for covariates were imputed by multiple imputation.[5]

References

1. Chavent M, Kuentz-Simonet V, Saracco J. Orthogonal rotation in PCAMIX. *Adv Data Anal Classif*. 2012;6(2):131-46.

2. Chavent M, Kuentz-Simonet V, Labenne A, Saracco J. Multivariate analysis of mixed data: The PCAmixdata R package, arXiv:1411.4911

3. Rosenbaum PR and Rubin DB. Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. *J R Stat Soc Ser B Methodol*. 1983;45(2):212-21

4. Chang C-CH. Applications of the propensity score weighting method in psychogeriatric research: correcting selection bias and adjusting for confounders. *Int Psychogeriatr*. 2017;29(5):703-706. doi:10.1017/S1041610216002490

5. Choi J, Dekkers OM, le Cessie S. A comparison of different methods to handle missing data in the context of propensity score analysis. *Eur J Epidemiol*. 2019;34(1):23-36. doi:10.1007/s10654-018-0447-z

Figure e-1. Scree plot of eigenvalues and percentages of explained variance associated to the four dimensions of the factorial analysis of mixed data on the three MRI biomarkers of cerebrovascular health, the Three-City Dijon study, 1999-2000 (n=1,623)
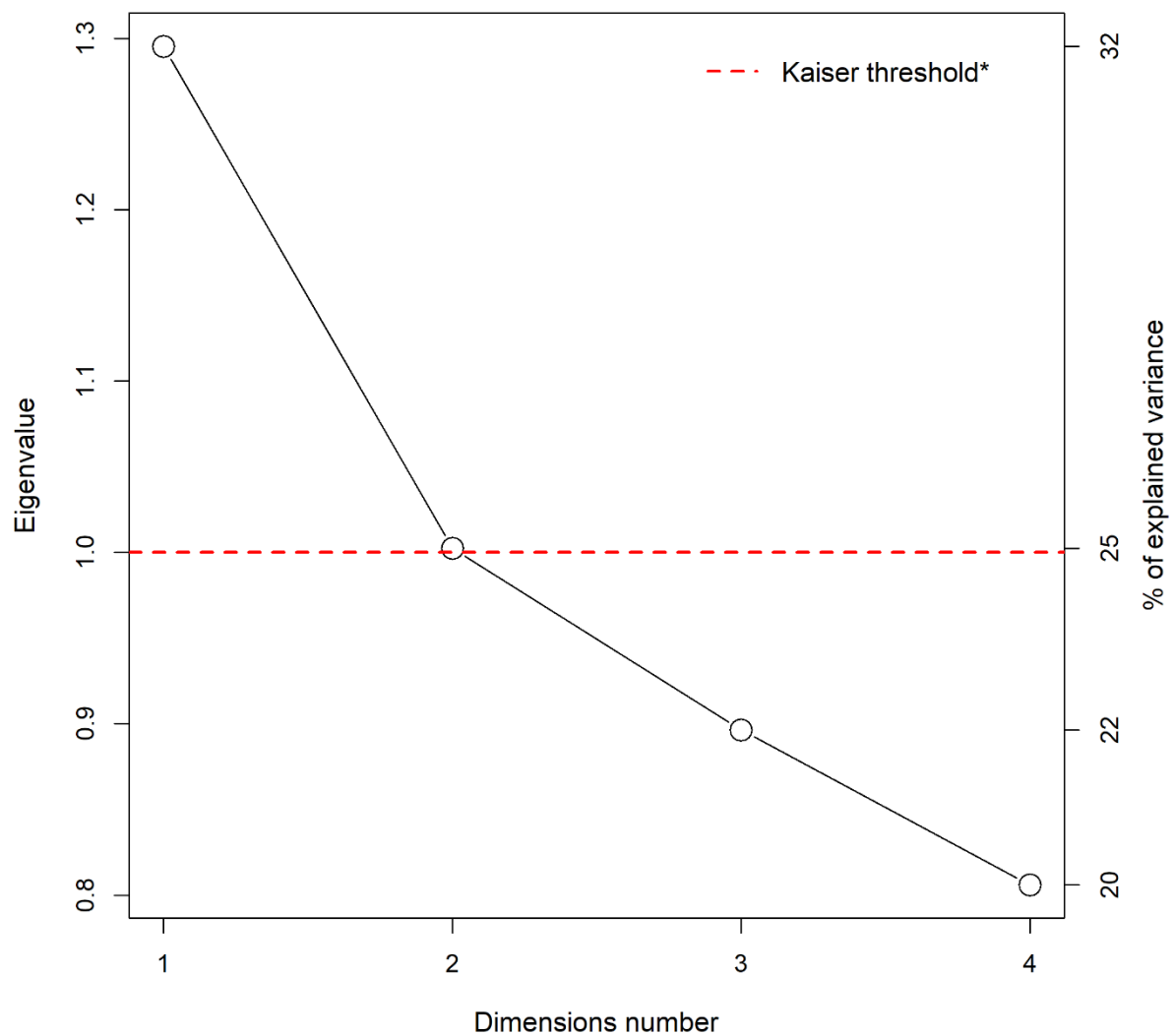
* Kaiser criterion (Kaiser. *Educational and Psychological Measurement*. 1960) proposes to retain only the dimensions with an eigenvalue greater than one; that is, the dimensions with an explained variance greater than the average (i.e. greater than 25% in our case where there is a maximum of 4 dimensions).

Table e-1. Comparison of characteristics of participants included in the analysis to those excluded, the 3C Dijon study, 1999-2000

	Total 3C Dijon population (n = 4,931)	Participants with MRI examination and fish intake data		
		Included (n = 1,623)	Excluded [a] (n = 301)	*P*-value [b]
Age (y), mean (SD)	74.6 (5.7)	72.3 (4.1)	73.5 (4.4)	<0.001
Female, n (%)	3,043 (61.7)	1,020 (62.8)	137 (45.5)	<0.001
Educational level, n (%)				0.58
None or primary	1,082 (22.0)	263 (16.2)	58 (19.3)	
Secondary	2,084 (42.3)	723 (44.6)	125 (41.7)	
High school	919 (18.7)	305 (18.8)	56 (18.7)	
University	839 (17.0)	330 (20.4)	61 (20.3)	
Monthly income (euros), n (%)				0.42
<750	217 (4.7)	48 (3.1)	9 (3.2)	
[750 – 1500[	1,466 (31.5)	448 (28.7)	97 (34.3)	
[1500 – 2250[	1,330 (28.6)	449 (28.8)	72 (25.4)	
≥ 2250	1,504 (32.4)	584 (37.5)	99 (35.0)	
Refused to answer	131 (2.8)	30 (1.9)	6 (2.1)	
Moderate to vigorous physical activity, n (%)	3,278 (75.3)	1,219 (79.7)	193 (68.9)	<0.001
Smoking, packs/year, mean (SD)	8.3 (16.8)	7.4 (15.9)	13.6 (19.8)	<0.001
Alcohol, drinks/week, mean (SD)	8.7 (10.3)	8.7 (10.1)	10.7 (11.3)	0.002
BMI (kg/m²), mean (SD)	25.6 (4.1)	25.3 (3.8)	26.5 (4.4)	<0.001
Diabetes, n (%)	459 (9.9)	125 (7.8)	42 (14.0)	0.001
Hypertension, n (%)	3,937 (80.0)	1,244 (76.6)	241 (80.1)	0.22
Hypercholesterolemia, n (%)	2,765 (58.9)	901 (55.9)	188 (62.9)	0.03
History of cardiovascular disease,[c] n (%)	1,416 (30.2)	327 (21.3)	144 (49.5)	<0.001
Antithrombotic treatment, n (%)	1,077 (21.9)	186 (11.5)	151 (50.2)	<0.001
Carotid plaque at ultrasound, n (%)	1,745 (52.5)	725 (46.1)	176 (62.0)	<0.001
CCA-IMT (mm), mean (SD)	0.70 (0.11)	0.68 (0.11)	0.69 (0.11)	0.32
*APOEε*4 status, n (%)	977 (21.2)	354 (21.8)	72 (24.0)	0.49
MMSE score (range, 0-30), mean (SD)	27.2 (2.1)	27.7 (1.8)	27.2 (2.0)	<0.001
BVRT score (range, 0-15), mean (SD)	11.2 (2.1)	11.8 (1.9)	11.5 (2.1)	0.01
TMT-A score, mean (SD)	28.5 (10.2)	30.5 (10.0)	28.7 (9.8)	0.004
TMT-B score, mean (SD)	12.9 (7.1)	14.3 (7.1)	12.5 (7.2)	<0.001
IST score, mean (SD)	33.0 (7.1)	34.2 (6.7)	32.9 (7.0)	0.001
Fish (times/week), mean (SD)	1.8 (1.1)	1.9 (1.1)	1.9 (1.2)	0.43
Meat (times/week), mean (SD)	4.8 (1.9)	4.8 (1.9)	4.7 (1.9)	0.48
Fruits and vegetables (times/week), mean (SD)	17.0 (4.1)	17.4 (3.9)	16.8 (4.1)	0.02
Legumes (times/week), mean (SD)	0.61 (0.56)	0.62 (0.56)	0.64 (0.64)	0.61
Olive Oil (preferred source of added fat), n (%)	1,813 (40.8)	619 (42.1)	98 (36.3)	0.09

Abbreviations: 3C = Three-City; *APOEε*4 = ε4 allele of the apolipoprotein E gene; BMI = body mass index; BVRT = Benton Visual Retention Test; CCA-IMT = common carotid artery intima-media thickness; IST = Isaacs Set Test; MMSE = Mini-Mental State Examination; SD = standard deviation; TMT-A/B = Trail Making Test part A/B.

Means and percentages are of non-missing values.

[a] We excluded participants with prevalent dementia (n=8), with history of stroke (n=83), with reported hospitalization for cardiovascular disease (i.e., myocardial infarction, and cardiac, abdominal aortic, carotid, coronary or leg artery surgery; n=87), with major acquisition artifacts on MRI scans (n=9) or brain tumors (n=7), and with missing data for at least one of the studied MRI biomarkers (n=107).

[b] *P*-values from Chi-square test for categorical variables and Student test for continuous variables.

[c] Non-hospitalized angina pectoris, cardiac rate disorder (including arterial fibrillation), heart failure or lower limbs arteritis.

Table e-2. Participants' characteristics by quartiles of global cerebrovascular disease burden indicator, the 3C Dijon study, 1999-2000 (n = 1,623)

|	Quartiles of global CVD burden indicator				P for trend [a]
	Q1	Q2	Q1	Q4	
Age (y), mean (SD)	71.5 (4.2)	72.0 (3.9)	72.3 (4.0)	73.4 (3.9)	<0.001
Female, n (%)	292 (71.9)	253 (62.3)	243 (60.0)	232 (57.1)	<0.001
Educational level (≤ secondary), n (%)	250 (61.6)	242 (59.9)	249 (61.5)	245 (60.3)	0.80
Monthly income (<1500 euros), n (%)	112 (29.2)	126 (32.8)	125 (33.0)	133 (34.8)	0.15
Moderate to vigorous physical activity, n (%)	316 (80.8)	314 (81.1)	300 (80.4)	289 (76.3)	0.06
Smoking, packs/y, mean (SD)	5.7 (12.7)	6.8 (14.9)	8.1 (18.4)	8.9 (17.0)	0.01
Alcohol, drinks /w, mean (SD)	8.4 (9.3)	8.0 (9.7)	8.7 (10.4)	9.6 (10.8)	0.03
BMI (kg/m²), mean (SD)	24.7 (3.6)	25.3 (3.6)	25.5 (3.8)	25.9 (3.8)	<0.001
Diabetes, n (%)	19 (4.7)	36 (9.0)	34 (8.5)	36 (9.0)	0.11
Hypertension, n (%)	276 (68.0)	309 (76.1)	315 (77.8)	344 (84.7)	<0.001
Hypercholesterolemia, n (%)	246 (61.2)	222 (54.8)	217 (53.8)	216 (53.6)	0.10
History of cardiovascular disease,[b] n (%)	86 (21.9)	69 (18.2)	75 (19.6)	97 (25.6)	0.05
Antithrombotic treatment, n (%)	39 (9.6)	35 (8.6)	40 (9.9)	72 (17.7)	<0.001
Carotid plaque at ultrasound, n (%)	168 (42.0)	169 (43.0)	176 (45.2)	212 (54.2)	<0.001
CCA-IMT (mm), mean (SD)	0.67 (0.10)	0.68 (0.11)	0.68 (0.11)	0.70 (0.11)	0.003
*APOE*ε4 status, n (%)	99 (24.6)	87 (21.6)	84 (20.9)	84 (20.8)	0.33
MMSE score (range, 0-30), mean (SD)	27.8 (1.8)	27.8 (1.7)	27.7 (1.8)	27.5 (1.9)	0.002
BVRT score (range, 0-15), mean (SD)	11.9 (1.9)	11.8 (1.8)	11.8 (1.9)	11.6 (2.0)	0.02
TMT-A score, mean (SD)	31.2 (9.8)	31.6 (10.2)	30.5 (10.3)	28.7 (9.6)	<0.001
TMT-B score, mean (SD)	14.6 (7.1)	15.0 (7.0)	14.1 (7.2)	13.4 (7.0)	0.005
IST score, mean (SD)	34.7 (6.7)	34.1 (6.8)	34.6 (6.5)	33.5 (6.9)	0.01
Fish (times/week), mean (SD)	2.0 (1.2)	1.9 (1.2)	1.8 (1.1)	1.7 (1.1)	0.001
Meat (times/week), mean (SD)	4.7 (1.8)	4.8 (1.9)	4.7 (1.9)	4.8 (1.9)	0.22
Fruits and vegetables (times/week), mean (SD)	17.4 (3.8)	17.6 (3.7)	17.4 (4.1)	17.1 (4.0)	0.17
Legumes (times/week), mean (SD)	0.63 (0.57)	0.61 (0.56)	0.63 (0.52)	0.62 (0.58)	0.99
Olive Oil (preferred source of added fat), n (%)	158 (42.4)	164 (44.3)	151 (41.0)	146 (40.7)	0.50

Abbreviations: 3C = Three-City; *APOE*ε4 = ε4 allele of the apolipoprotein E gene; BMI = body mass index; BVRT = Benton Visual Retention Test; CCA-IMT = common carotid artery intima-media thickness; CVD = cerebrovascular disease; IST = Isaacs Set Test; MMSE = Mini-Mental State Examination; SD = standard deviation; TMT-A/B = Trail Making Test part A/B.

Means and percentages are of non-missing values. Missing values (by decreasing %): 9.4% for olive oil consumption, 7.1% for alcohol consumption, 5.7% for physical activity, 5.1% for non-hospitalized cardiovascular disease, 4.4% for CCA-IMT, 4.3% for TMT-B, 3.9% for monthly income, 3.1% for carotid plaque, 1.5% for TMT-A, 1.4% for tobacco consumption, 1.0% for diabetes, 0.8% for *APOE*ε4 status and BVRT score, 0.6% for hypercholesterolemia, 0.5% for IST score, 0.1% for educational level, BMI, MMSE and fruits and vegetables intake.

[a] P-values for linear trends across quartiles obtained from linear regressions for continuous variables and logistic regressions for categorical variables, using a continuous independent variable for CVD indicator score in which participants in a given quartile were assigned the median value.

[b] Non-hospitalized angina pectoris, cardiac rate disorder (including arterial fibrillation), heart failure or lower limbs arteritis (history of hospitalized cardiovascular diseases were excluded from study sample).

Figure e-2. Associations of fruit and vegetable, legume, meat and olive oil intakes with the global cerebrovascular disease burden indicator, the 3C Dijon study, 1999-2000 (n=1,623)
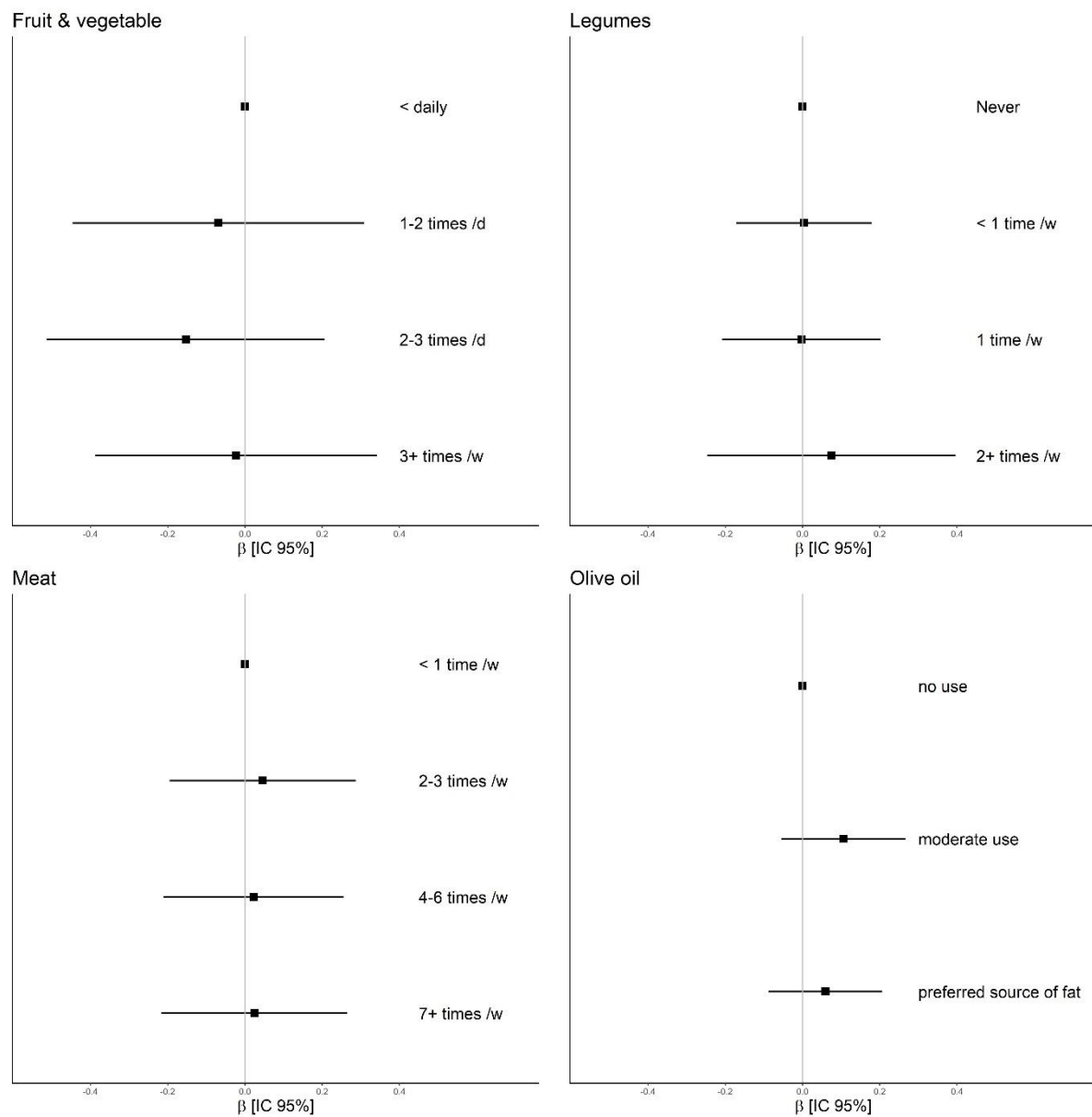
β coefficients and 95% confident intervals (CIs) for fruit and vegetable intake, legume, meat and olive oil intakes estimated by a mutually-adjusted linear regression model controlled for: fish intake (primary exposure), total intracranial volume, age, sex, educational level, monthly income, body mass index, tobacco consumption, alcohol consumption, engagement in moderate to vigorous physical activity, diabetes, history of non-hospitalized cardiovascular diseases, hypertension, hypercholesterolemia and carrying ε4 allele of the apolipoprotein E gene. Interactions of dietary intakes and age on CVD burden indicator were not statistically significant (*P* >0.05 for interaction terms).

Figure e-3. Association of fish intake with the global cerebrovascular disease burden indicator (propensity score weighed analysis accounting for non-inclusion), the 3C Dijon study, 1999-2000 (n = 1,623)
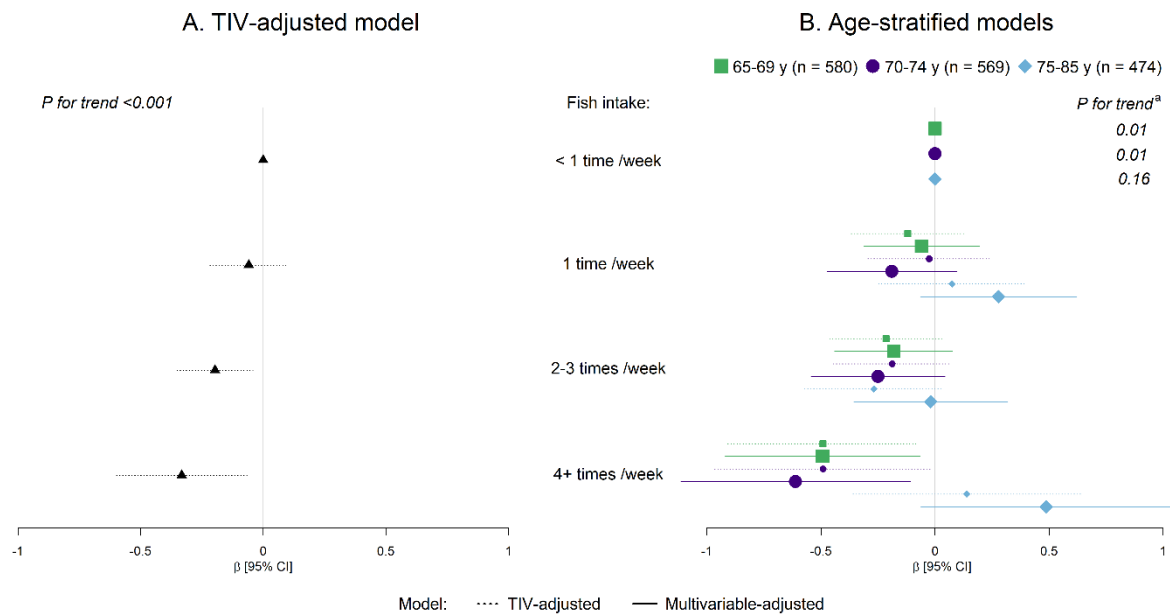
β coefficient and 95% confident intervals were estimated by linear regression models with fish intake as a main explanatory variable (upper intake categories versus <1 time per week as a reference). Regression models were weighted by the inverse of the propensity score derived from logistic regression with inclusion/exclusion status as the outcome. Panel A: model adjusted for total intracranial volume (TIV). Panel B: age-stratified models, (i) adjusted for TIV (dotted lines) and (ii) further adjusted for sex, educational level, monthly income, body mass index, tobacco consumption, alcohol consumption, intakes of fruits and vegetables, legumes, meat and olive oil, engagement in moderate to vigorous physical activity, diabetes, history of non-hospitalized cardiovascular diseases, hypertension, hypercholesterolemia and carrying ε4 allele of the apolipoprotein E gene (solid lines).

P-value for linear trend across categories was obtained using a continuous variable in which participants in a given category were assigned the median number of times of fish intake per week.

[a] *P*-values for linear trends across fish intake categories estimated in age-stratified TIV-adjusted model (*P* for trend in multivariable-adjusted models were: 0.01 in 65-69 years; 0.03 in 70-75 years; and 0.95 in ≥75 years).