Polygenic Risk for Alzheimer's Disease in Healthy Aging: Age-related and APOE-driven Effects on Brain Structures and Cognition

Method S1: Variable definitions and detailed exclusion criteria

Confounding variables

Month (data field 52) and year (data field 34) of birth (Note the exact day of birth was unavailable due to UKB policy, so we set it to the first day of the month for all participants), sex (data field 31), the index of deprivation (data field 477), and education score (data field 18657) were obtained from baseline characteristics at recruitment. Scanning date (data field 99) and scanning center (data field 103) were considered to mitigate the drifting effects and center effects. Genetic principal components (data field 22009) were considered to control for population stratification. Two vascular diseases including hypertension and diabetes were considered for their impacts on the nervous system. Nurses inquired about participants' medical histories, including any self-reported diagnoses. The status of hypertensive and diabetic were based on self-reported diagnoses, the International Classification of Disease 10th version (ICD-10) code (data field 41270), and body biological measures (**Table S1.1**)

Hypertension	Meets any following conditions • Ever diagnosed with hypertension (6150) • ICD10 hypertension diagnosis (41720) [a] • Blood pressure medication (6153) • Systolic blood pressure ≥ 140mmHg (4080) [b] • Diastolic blood pressure ≥ 90mmHg (4079) [b]	Pulse pressure (mmHg), defined as the difference between systolic pressure (4080) and diastolic blood pressure (4079) [c]
Diabetes	Meets any following conditions • Ever diagnosed with diabetes (2443) • ICD10 diabetes diagnosis (41720) [c]	HbA1c level (30750)

- Insulin medication (6153)
- Glucose level ≥ 7 mmol/L (30740)
- HbA1c level ≥ 48 mmol/mol (30750)

The number within each parenthesis is related data filed ID of UK Biobank.

[a] Alcohol ICD-10 code for hypertension: 'I10', 'I11', 'I110', 'I119', 'I12', 'I120', 'I12', 'I13', 'I130', 'I131', 'I132', 'I139', 'I15', 'I150', 'I151', 'I152', 'I158', 'I159'.

[b] Both systolic and diastolic blood pressures were calculated as the mean of two corresponding automated measures.

[c] ICD-10 code for diabetes: 'E10', 'E100', 'E101', 'E102', 'E103', 'E104', 'E105', 'E106', 'E107', 'E108', 'E109', 'E11', 'E110', 'E111', 'E112', 'E113', 'E114', 'E115', 'E116', 'E117', 'E118', 'E119', 'E12', 'E120', 'E121', 'E122', 'E123', 'E124', 'E125', 'E126', 'E127', 'E128', 'E129', 'E13', 'E130', 'E131', 'E132', 'E133', 'E134', 'E135', 'E136', 'E137', 'E138', 'E139', 'E14', 'E140', 'E141', 'E142', 'E143', 'E144', 'E145', 'E146', 'E147', 'E148', 'E149'.

Table S1.1 Definition of hypertension and diabetes

Exclusion criteria

Based on self-reported non-cancer illness (data field 20002) and (data field 41270), we excluded participants with stroke, including brain hemorrhage, transient ischemic attack, subdural hemorrhage/haematoma, subarachnoid hemorrhage, and ischemic stroke, along with cerebral aneurysm, infections of the nervous system such as brain abscesses/intracranial abscesses, encephalitis, and meningitis. Additionally, chronic/degenerative neurological problems like motor neuron disease, myasthenia gravis, multiple sclerosis, Parkinson's disease, dementia (including Alzheimer's disease and cognitive impairment), other demyelinating diseases (excluding multiple sclerosis), cerebral palsy, meningioma/benign meningeal tumors, benign neuromas, neurological injuries/trauma, head injuries and other severe brain diseases **(Additional file 2: Table S1)**.

Method S2: Neuropsychological testing and cognitive factors

Considered cognitive tests

We considered eight tests with robust test-retest reliability [1] from the UK Biobank cognitive test batteries including the reaction time test (data field 20023), numeric memory test (data field 4282), fluid intelligence (data field 20016), two trail-making tests (TMT-A: data field 6348; TMT-B: data field 6350), pairs matching test (data field 521), matrix pattern completion test (data field 6373), symbol digit substitution test (data field 23324) and tower-rearrange test (data field 21004). The description and related cognitive domains are summarized in **Additional file 2: Table S2**. Abnormal records for reaction time (less than 50ms or greater than 2000ms as during which nothing present on screen) and numeric memory test (-1) were set as missing values.

Random forest-based imputation

On the observation of the missing pattern of missing at random (MAR) of the cognitive tests (**Fig S2.1**), we utilized the *mice* package to perform random forest (RF) imputation, which is an ensemble learning method that combines multiple decision trees to make predictions for both continuous and discrete missing data [2]. To achieve a reasonable imputation, Our implementation involved 100 multiple imputations and 200 iterations on all participants from imaging visit phase 2 with at least one cognitive performance available (N = 52,327). We evaluated the imputation results separately for discrete and continuous variables. For discrete variables such as ‘fluid intelligence', ‘numeric memory', ‘symbol digit', ‘tower rearrange', and ‘matrix pattern', we conducted chi-square tests to compare the proportions of each value between the original and imputed data. For continuous variables such as ‘TMT-A', ‘TMT-B', and ‘reaction-time', we performed Levene tests to compare the original and imputed data in terms of mean and variance, with the significance level set at

Bonferroni corrected 0.05. No significant difference was observed on the distribution was observed before and after the imputation (p_{Bonf}>0.05, **Table S2.1**).

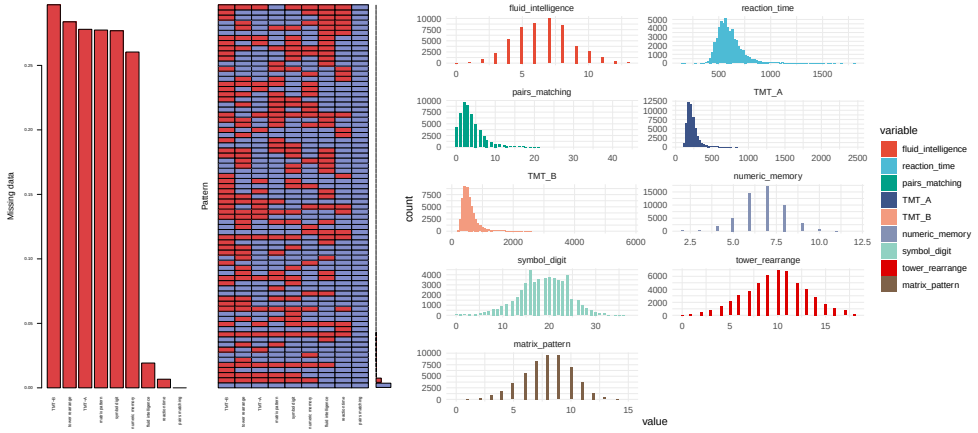

Fig. S2.1 Missing patterns and distribution of the imputed cognitive data

Test	Statistic	P value	DF	Median (raw)	Median (imputed)	P_{bonf}
TMT_A	3.21690654	0.07288453	1	212	211	0.21865359
TMT_B	0.00393531	0.94998002	1	523	522	1
Reaction time	0.00117823	0.97261781	1	581	581	1
Fluid intelligence	182	0.23398264	169	7	7	1
Numeric memory	110	0.23220478	100	7	7	1
Symbol digit	1368	0.24065802	1332	19	19	1
Tower rearrange	342	0.23561022	324	10	10	1
Matrix pattern completion	240	0.23476617	225	8	8	1

Table S2.1 Cognitive data before and after the imputation

Cognitive factor extraction

The imputed cognitive data underwent tests for factor analysis suitability using Bartlett's test and the Kaiser-Meyer-Olkin criterion. The significant result from Bartlett's test of sphericity ($\chi^2(28) = 80164.92$, $p < .001$) as well we KMO indexes (overall KMO = 0.863) suggest high adequacy for factor analysis across the selected tests. We extract 4 cognitive factors based on the potential cognitive field measured with each test (**Additional file 2: Table S2**). Briefly, executive function (**EF**) encapsulates cognitive tasks associated with higher-order processing, task-switching, and problem-solving. Memory function indicate a concentration on memory-related abilities and general cognitive aptitude. Visuospatial

Function (**VS**) suggest prowess in spatial cognition and visual problem-solving abilities. Reasoning function (**RF**) measures the ability to problem-solving with logics.

Confirmatory factor analysis, with the *cfa* function from the *lavaan* package [3], was then conducted to extract cognitive factors corresponding to the related domains. The CFA model suggested a fairly good fit to the data (RMSEA=5.81e-03(90%CI[3.41e-03,8.31e-03]), CFI=1, TLI=0.999, $\chi^2(11)=30.44$, p=1.349e-03). All extracted cognitive factors showed robust sensitivity to aging ($p < 10^{-16}$).

References

1. Fawns-Ritchie C, Deary IJ. Reliability and validity of the UK Biobank cognitive tests. PLoS One. 2020;15:e0231627.
2. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45:1–67.
3. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software. 2012;48:1–36.