Aging Brain 3 (2023) 100081

Contents lists available at ScienceDirect

Aging Brain

journal homepage: www.elsevier.com/locate/nbas

Ranking the risk factors for Alzheimer's disease; findings from the UK Biobank study

Michael Allwright [a], Hamish D Mundell [b], Andrew N McCorkindale [b], Richard I. Lindley [c], Paul J. Austin [a], Boris Guennewig [a,1], Greg T Sutherland [b,1,*]

[a] Brain and Mind Centre and School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, NSW 2006, Australia
[b] Charles Perkins Centre and School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, NSW 2006, Australia
[c] Westmead Applied Research Centre, Sydney Medical School, University of Sydney, NSW 2006 and George Institute for Global Health, Newtown, NSW 2042, Australia

ARTICLE INFO

Article history:
Received 28 December 2022
Revised 25 May 2023
Accepted 6 June 2023

Keywords:
Alzheimer's disease
UK Biobank
Machine learning
Apolipoprotein E
Liver
Multimorbidity

ABSTRACT

Background: The cause of the most common form of dementia, sporadic Alzheimer's disease (AD), remains unknown. This may reflect insufficiently powered studies to date for this multi-factorial disorder. The UK Biobank dataset presents a unique opportunity to rank known risk factors and determine novel variables.
Methods: A custom machine learning approach for high dimensionality data was applied to explore prospectively associations between AD in a sub-cohort of 156,209 UK Biobank participants aged 60–70 including more than 2,090 who were subsequently diagnosed with AD.
Results: After the possession of the APOE4 allele, the next highest ranked risk factors were other genetic variants within the TOMM40-APOE-APOC1 locus. When stratified by their apolipoprotein *epsilon* 4 (APOE4) carrier status, the most prominent risk factors in carriers were AST:ALT ratio, the "number of treatments/ medications" taken as well as "time spent in hospital" while protection was conferred by "Sleeplessness/Insomnia". In non-APOE carriers, lower socioeconomic status and fewer years of education were highly ranked but effect sizes were small relative to APOE4 carriers.
Conclusions: Possession of the APOE4 allele was confirmed as the most important risk factor in AD. Other TOMM40-APOE-APOC1 locus variants further moderate the risk of AD in APOE4 carriers. Liver pathology is a novel risk factor in APOE4 carriers while "Sleeplessness/Insomnia" is protective in AD irrespective of APOE4 status. Other factors such as "Number of treatments/ medications" suggest that multimorbidity is an important risk factor for AD. Future treatments aimed at co-morbidities, including liver disease, may concomitantly lower the risk of sporadic AD.

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia affecting over 50 million people worldwide [1] and with no proven disease-modifying therapies, it presents major social, economic and policy challenges.

The common, sporadic forms of AD are generally thought to be caused by a complex interaction of genetic and environmental factors [2], although this hasn't been empirically tested to date due to a lack of study power. A 2020 Lancet Commission proposed a set of 'consensus associations' with dementia based on known epidemiolog-

* Corresponding author.
E-mail address: g.sutherland@sydney.edu.au (G.T Sutherland).
[1] Both lead authors contributed equally.

https://doi.org/10.1016/j.nbas.2023.100081
2589-9589/

ical data including age, gender, diabetes, hypertension, smoking, traumatic brain injury, excessive alcohol consumption, pollution, and apolipoprotein E (APOE4) status [3]. More than 50 other genetic loci and numerous environmental factors such as mild traumatic brain injury [4] and periodontitis [5] have also been associated with AD, but many others may be unknown.

The UK Biobank (UKB) is the largest, most densely phenotyped, longitudinal study in the world, featuring over 500,000 participants. It has so far enabled several specific associations with dementia to be tested, including concomitant cancer [6], frailty [7], meat consumption [8], pollution [9], cognition [10], and cardiovascular risk factors [11,12]. In each of these studies, variables associated with the hypothesis in question were modelled alongside confounders such as age, gender, and the presence of the APOE4 allele. One study selected 30 potential risk factors and modelled these using logistic and lasso regression to develop a novel dementia risk score (preprint only, [13]). However, the full range of variables available in the UKB have not yet been modelled to give an unbiased ranking of their importance in predicting AD.

Machine learning has developed substantially over the last 20 years. Non-parametric tree-based solutions are effective at modelling the associations, interactions and non-linear relationships between thousands of independent variables and a dependent variable, scoring the "feature importance" of individual variables as a ranking of their contribution to the accuracy of a given model. EXtreme Gradient Boosting (XGBoost) is a leading algorithm [14], but one of the challenges with XGBoost and similar algorithms are their "black-box" nature and the difficulty of determining a ranking of variables responsible for the model's prediction. SHapley Additive exPlanations (SHAP) [15] is a feature importance methodology which employs co-operative game theory to interpret outputs for classification and regression algorithms, enabling a ranked list of feature importance to be developed for any given model. This algorithmic combination has been applied to the UKB to predict myocardial infarction [16].

Here, for the first time, a combination of XGBoost and SHAP was applied to a combination of health-related questionnaire data, longitudinal inpatient data (ICD10), blood assays, cognitive testing, and genetic data from the UKB allowing simultaneous consideration of over 1000 potential risk factors for AD. The aim was to provide a ranked set of risk factors for AD that would focus future mechanistic studies towards new therapeutics.

2. Materials and methods

2.1. Sample selection

The UKB study recruited 502,253 subjects, aged 37–73 years in the United Kingdom between 2006 and 2010. A raft of clinical measurements and assays were performed during the initial attendance at the assessment centre ("baseline"), including clinical pathology screens, genotyping, neuroimaging, and cognitive testing as well as health records and self-reported demographic and wellness data [17]. Participants aged 60–70 years who had genotype information available at baseline were selected. Exclusion criteria included those already diagnosed with dementia at baseline or who developed dementia within 2 years of baseline; died of something other than AD or dementia within 10 years of baseline or had been diagnosed with Parkinson's disease prior to baseline. This left a total of 156,209 participants, of which 2,090 had developed AD within an average of 8.2 years from baseline (Fig. 1).

Genotype information derived using the UKB Axiom array platform for each participant was processed using Plink 1.9 [18]. A single variable "APOE4 carrier" was computed from the two single nucleotide polymorphism (SNPs), rs429358 and rs7412 (haplotype) data. A score of zero was given to for those with no APOE4 alleles, one for those with one allele and two for APOE4 homozygotes. Otherwise for genetics, and to reduce dimensionality, 38 AD-related SNPs from the most recent and largest genome-wide association data were included in the models here [19].

The label "AD" was given to those participants who were diagnosed between two and ten years of their baseline visit to the UKB assessment centre with one of the following ICD10 codes, relating to AD (G30.0, G30.1, G30.8 or G30.9). The most recent inpatient data (release date: September 2021) was used to identify all ICD10 codes corresponding to any condition for which the number of cases across our cohort exceeded 5,000 (for subsequent test validity) and for which a diagnosis was received prior to attendance at the assessment centre. This resulted in binary features corresponding to a participant having been diagnosed with (1) or not (0) any given disease. All variables with<20% missing observations were selected for the subsequent analysis. "One hot encoding" of categorical variables and imputation of missing values using miceforest [20] was applied to structure the data for XGBoost. This resulted in a full set of 1,002 candidate features (V_S) (Fig. 1).

A set of variables relating to "known associations" from a recent consensus report by Livingston and colleagues were computed for a comparison model [21]. This set included variables related to traumatic brain injury, age, gender, APOE4 carriers, hearing difficulties, pollution, hypertension at baseline, diabetes at baseline, smoker status, body mass index, qualifications, frequency of friend and family visits (as a proxy for loneliness) and IPAQ (Physical Activity Questionnaire) activity group. These variables comprised the "known associations" variable set, denoted V_K.

2.2. Machine learning

2.2.1. Data processing and model selection

To avoid data leakage, data was first split into training data (D_T) and a holdout dataset (D_V), with D_V containing a random sample of 30% of AD cases and 30% of controls with the remainder of each being in D_T. Three classification models were applied to D_T using the full variable set V_S. These were evaluated using nested cross-fold validation with three folds (two training and one validation), resampled 50 times. The mean Area Under Curve – Receiver

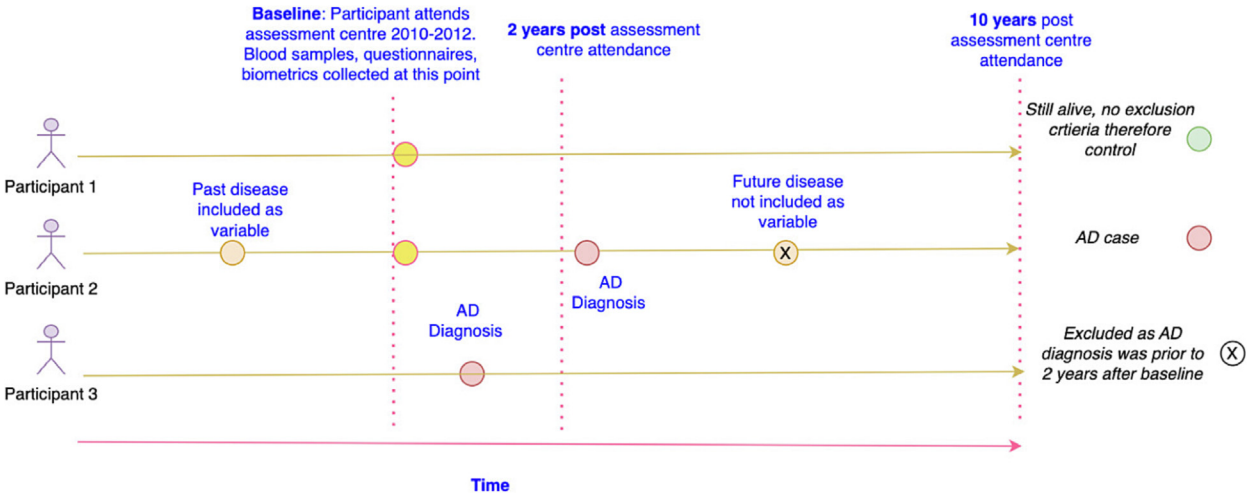

Fig. 1. UK Biobank timeline and case/control inclusion/exclusion criteria. A schematic demonstrates our selection criteria for hypothetical UKB participants. In this case, Participant 1 would be selected in our analysis as they attended the assessment centre and did not meet any of the exclusion criteria, nor were they ever diagnosed with AD. Participant 2 is included as a case, with a past disease coded as an independent variable as it was diagnosed prior to their attendance at the assessment centre, and they were diagnosed with AD over 2 years after their attendance. Participant 3 is excluded as they received an AD diagnosis prior to 2 years after attending the baseline assessment centre.

Operating Characteristics Curve (AUC) of each model trained on a training fold and evaluated on the validation fold was used to determine the best performing classifier. XGBoost (0.77) outperformed random forest (0.75) and Support Vector Machines (0.74) and hence was selected for the subsequent analysis.

Hyper-parameter optimization was performed using a grid search [22] on the following XGBoost hyperparameters: "minimum child weight", "gamma," "subsample rate", "maximum depth" and scale positive weight. The set of hyperparameters which generated the highest AUC values for the model were selected – these were minimum child weight = 5, gamma = 2, subsample rate = 1, maximum depth = 5 and scale positive weight = 1.

2.2.2. Model pipeline

Due to the overwhelming importance of age in predicting AD, the data were age matched first such that the case to control ratio was the same for all ages (in years) at baseline. Because the number of controls far exceeded the number of cases, we resampled controls using Monte Carlo cross-fold validation [23]. In this way, controls from D_T were resampled 20 times for the same set of cases to create 20 age matched training datasets $d_T \subset D_T$. The hyperparameter tuned XGBoost model was trained on each d_T, and the corresponding AUC and mean SHAP score for each variable in V_S was evaluated using validation dataset D_V for each resample. This approach was repeated with D_T restricted to the set of known risk factors (V_K) for dementia. A *t*-test was performed to compare the mean AUCs across all resamples in each case. A new ordered list of variables (risk and protective factors) was determined by selecting the top 20 variables $V_T \subset V_S$, ordered by the mean SHAP score across all resamples. The complete process was then repeated for APOE4 carriers and non-APOE4 carriers separately.

AD incidence [24] is defined as number of new cases within 2–10 years post baseline divided by the total eligible UKB population at 2 years post baseline. AD Incidence was calculated for each value for variables in V_T, and the top output variables reported for each cohort: APOE4 carriers, non-APOE4 carriers and both. The analysis was performed using the validation dataset, D_V to avoid overfitting. Chi-squared proportionality tests [25] were used to determine if there was a statistically significant difference each grouping of each variable and all other groupings. Continuous variables were split into 'high' (greater than the variable's median value across each cohort) and 'low' (below the variable's median value across each cohort). A Benjamini Hochberg correction [26] was performed on the resulting p values to correct for multiple comparisons. All p values presented are the adjusted p values.

All analyses were performed in Python (3.10.0) and the GitHub codebase is available at https://github.com/binfn-stats/ukb-dementia-shap/tree/main.

3. Results

A cohort of 156,209 UKB participants, aged 60–70 years-of-age at baseline met the inclusion criteria here. 2,090 participants had been diagnosed with AD between two and 10 years from baseline, mean 8.2 years (Table 1). The cohort was predominantly White (96%) and there were more female controls (55%) than males (see Supplementary file S1 for full demographic information on UKB participants).

In terms of classification accuracy, the full UKB model (V1) (AUC = 0.77), outperformed a model made up of known (consensus) risk factors (V2; AUC = 0.67 (3)) (Table 2). Focussing purely on APOE4 carriers, the UKB model performance remained high (AUC = 0.75), compared

Table 1
UKB participant demographics.

	AD Cases	Controls
Mean Age (Years)	65.6 +/- 2.6	63.9 +/- 2.8
Gender		
Male	1.006	69,060
Female	1,084	85,059
Total	2,090	154,119
Ethnicity		
African	11	459
Asian	31	2,374
White	1,993	148,380
Other	42	2,130
Unknown	13	776
Total	**2,090**	**154,119**

Table 2
Performance of known risk factors versus the UKB model.

APOE4	Known Risk Factors	UKB Model
All	0.67+/-0.01	0.77+/-0.004***
Carriers	0.55+/-0.01	0.75+/-0.01***
Non-carriers	0.56+/-0.01	0.64+/-0.02***

Values are Receiver Operator Curve - Area Under Curve (AUC). AUC definitions: *** p < 0.001.

to the known risk factors (AUC = 0.55) (Table 2). For non-APOE4 carriers the performance of the UKB model was lower (AUC = 0.64) but remained better than known risk factors (AUC = 0.56) (Table 2; see Supplementary Fig. 1 for the ROCs).

In the UKB model, the number of APOE4 alleles (represented by the SNP, rs429358; the alternative 'C' variant) had the highest mean SHAP score followed by other variants within the TOMM40-APOE-APOC1 locus on chromosome 19 (Fig. 2A). rs4420638 is 343 bp 3′ of *APOC1*, rs6857 is in the 10th and final exon of *NECTIN2,* the gene immediately (45 Kb) upstream of *TOMM40* and rs769449 in is intron 4 of the *APOE* gene itself. There was a relative decrease in effect sizes for the remaining factors, with "Sleeplessness/Insomnia" conferring a protective effect and high serum AST:ALT ratio the next highest risk factor (Fig. 2A). The presence of both "Number of treatments/medications taken" and "Spells in hospital" in the top 20 ranked factors, appear consistent with multimorbidity or frailty being associated with AD cases.

Given the importance of the *APOE4 allele*, the cohort was stratified into carriers and non-carriers. It was hypothesized that in the absence of *APOE4* the effect size of some factors would increase, or novel factors be identified. Among the *APOE4* carriers, two copies of *APOE4* (rs429358) versus a single copy was the highest ranked factor (Fig. 2B), followed by AST:ALT ratio. rs7412, was protective here as the alternative (T) allele is associated with *APOE2* (if the common variant 'T' is also at rs429358). Contrary to our hypothesis, almost all risk factors for non-APOE carriers had lower mean SHAP scores with wide confidence intervals. The latter suggesting that these rankings are unstable and subject to variable rankings across repeated measures (Fig. 2C). The "North coordinate at birth" was the leading risk factor in non-APOE4 carriers with the spirometric variable "FEV1/ FVC ratio Z-score" while "urate" was protective. (Fig. 2C). "North coordinate at birth" (positive association) and "Average household income before tax" (negative association) suggest an inverse association between socioeconomic advantage and AD. Similarly, "Education score (England)", another prominent risk factor in the non-carriers, is a UKB measure of "*deprivation in terms of education, skills and training*". Supplementary file S1 contains a complete list of all factors modelled and their associated mean SHAP scores.

The impact of these factors on AD incidence was then explored in three cohorts: APOE4 carriers, non-APOE4 carriers and both to explore dose-dependent effect of the APOE4 allele. For both and APOE4 carriers rs429358 was the most prominent followed by other genetic variants within the TOMM40-APOE-APOC1 locus (Table 3). Using the National Cancer Institute's LD pair tool based on Phase 3 of the 1000 Genomes Project [27] rs4420638 (near APOC1; D' = 0.58), and rs6857 (*NECTIN2*; D' = 0.63) were both in modest linkage disequilibrium (LD) with rs429358 and their additional effects reflect nuances within the wider APOE4 haplotype [28]. rs769449 was in almost complete LD (intronic APOE; D' = 0.99) and likely reflects the APOE4 effect. In non-APOE4 carriers, possessing copies of rs4420638, was the most potent risk factor (accounting for 7.7% AD incidence) reinforcing the idea that other TOMM40-APOE-APOC1 locus SNPs are important in AD risk. The "Frequency of friend/family visits" and frequency of "Ability to confide" were also important variables.

4. Discussion

A ranking of the risk factors for AD was derived from the world's largest and most comprehensive longitudinal community study in combination with the state-of-the-art machine learning pipeline. At the time of this study, 2090 UKB participants, who attended the assessment centre between the age of 60 and 70 years, had subsequently developed AD 2 or more years after their visit. After possession of the APOE4 allele, the most potent risk factors were additional SNPs at the TOMM40-APOE-APOC1 locus and AST:ALT ratio while "Sleeplessness/Insomnia" was the major protective effect.

The relative effect size of the APOE4 haplotype here reiterates that understanding how this common variant mechanistically modifies AD risk is the most pressing need for the AD research community. The effect of APOE4 has been known since 1993 but the mechanism by which it modulates AD risk is unknown [29]. Generally, APOE binds lipids and delivers them to cells via receptor-mediated uptake [30]. In the brain the major receptor is lipoprotein receptor-related protein 1 (LRP1). During ageing, possession of APOE4 is associated with decreased hippocampal volume in females [31] and breakdown of the blood–brain-barrier [33]. APOE4 binds to Aβ with more affinity than APOE3 or APOE2 and this may account for the propensity in carriers to form Aβ plaques [31]. In AD, APOE4 is associated with both plaques and neurofibrillary tangles (insoluble tau) and appears to adversely affect cognitive function via these pathognomonic entities [32]. It is also

Fig. 2. SHAP scores and AD risk. Bar plots showing mean SHAP scores for the entire cohort (A), APOE4 carriers (B) and non-APOE4 carriers (C). The colour of the bar represents features which positively impacted on AD risk in the UKB model (red) and those with a negative or protective association (blue). Black bars indicate confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

associated with decreased CSF Aβ 1–42 [32]. A second mechanism is that APOE4 carriers have poorer Aβ clearance via APOE-dependent LRP1-mediated endocytosis within the brain [33,34] and across the BBB [35] and blood-CSF barriers [36]. Soluble LRP1 also binds the majority of Aβ in plasma, creating a peripheral 'sink' [37]). A third or even adjunctive mechanism maybe relatively less Aβ uptake by microglia of APOE4 carriers via a Triggering receptor expressed on myeloid cells-2 (TREM2)-APOE pathway [38,39]. Whilst TREM2 is a receptor expressed on all myeloid cells, it is restricted to microglia in the brain, and rare variants increase the risk of AD [40]. TREM2-dependent activated microglia appear to delay Aβ-driven tau-seeding [41] providing an opportunity whereby immune modulation could disarticulate Aβ accumulation from the neurodegenerative (tau) stages of AD.

Table 3
AD incidence in APOE carriers and non-APOE carriers.

Attribute	Value	All	Carriers	Non carriers
AST_ALT_ratio	High	0.0389 (***)	0.0626 (***)	0.0229 (***)
	Low	0.0277 (***)	0.0426 (***)	0.0188 (***)
Ability to confide	Never or almost never	0.0363 (**)	0.0546 (ns)	0.025 (***)
	Once every few months	0.0421 (***)	0.0647 (***)	0.0276 (***)
	About once a month	0.0328 (ns)	0.0524 (ns)	0.0199 (ns)
	About once a week	0.0336 (ns)	0.0563 (ns)	0.0186 (ns)
	2–4 times a week	0.0321 (ns)	0.0496 (ns)	0.0211 (ns)
	Almost daily	0.0316 (***)	0.0504 (**)	0.0194 (**)
Aspartate aminotransferase	High	0.036 (***)	0.0576 (***)	0.0216 (ns)
	Low	0.0307 (***)	0.0478 (***)	0.02 (ns)
Average household income before tax	<18,000	0.0396 (***)	0.0628 (***)	0.0245 (***)
	18,000 to 30,999	0.034 (ns)	0.0551 (*)	0.0199 (ns)
	31,000 to 51,999	0.0242 (***)	0.0337 (***)	0.0191 (ns)
	52,000 to 100,000	0.027 (***)	0.0435 (***)	0.0159 (**)
	Greater than 100,000	0.0245 (**)	0.0516 (ns)	0.0054 (***)
C-reactive Protein	High	0.0276 (***)	0.0429 (***)	0.021 (ns)
	Low	0.039 (***)	0.0623 (***)	0.0206 (ns)
Cooked Vegetable Intake	High	0.0354 (**)	0.0517 (ns)	0.0256 (***)
	Low	0.0327 (**)	0.0529 (ns)	0.0193 (***)
Drive faster than motorway speed limit	Do not drive on the motorway	0.0375 (***)	0.0584 (***)	0.0242 (***)
	Sometimes	0.0287 (***)	0.0462 (***)	0.0172 (***)
	Often	0.0248 (***)	0.0419 (**)	0.0131 (***)
	Most of the time	0.0063 (***)	0.0147 (***)	0.0 (***)
Duration to first press of snapbutton	High	0.0366 (***)	0.0579 (***)	0.0228 (***)
	Low	0.0301 (***)	0.0474 (***)	0.0189 (***)
Education score (England)	High	0.0377 (***)	0.0585 (***)	0.0244 (***)
	Low	0.029 (***)	0.0468 (***)	0.0173 (***)
Fev1 fvc ratio zscore	High	0.0365 (***)	0.0592 (***)	0.0214 (ns)
	Low	0.0302 (***)	0.046 (***)	0.0203 (ns)
Frequency of friend/ family visits	Never or almost never	0.0755 (***)	0.0833 (***)	0.0772 (***)
	Once every few months	0.035 (ns)	0.0573 (ns)	0.0207 (ns)
	About once a month	0.0334 (ns)	0.0565 (ns)	0.0182 (ns)
	About once a week	0.0309 (***)	0.0491 (**)	0.0189 (**)
	2–4 times a week	0.0348 (**)	0.0531 (ns)	0.0234 (***)
	Almost daily	0.0307 (*)	0.0522 (ns)	0.0158 (***)
Gamma Glutamyltransferase	High	0.0334 (ns)	0.0528 (ns)	0.0201 (ns)
	Low	0.0333 (ns)	0.0524 (ns)	0.0215 (ns)
Home location at assessment east coordinate	High	0.0347 (**)	0.056 (***)	0.0207 (ns)
	Low	0.032 (**)	0.0494 (***)	0.021 (ns)
Mean time to correctly identify matches	High	0.0342 (*)	0.054 (ns)	0.022 (*)
	Low	0.0324 (*)	0.0513 (ns)	0.0197 (*)
North Coordinate at birth	High	0.0388 (***)	0.0617 (***)	0.0238 (***)
	Low	0.0279 (***)	0.0436 (***)	0.0179 (***)
Number of treatments/medications	High	0.0431 (***)	0.0666 (***)	0.0278 (***)
	Low	0.0275 (***)	0.0442 (***)	0.0167 (***)
Particulate matter air pollution (pm10)	High	0.0371 (***)	0.0595 (***)	0.0216 (ns)
	Low	0.0296 (***)	0.0458 (***)	0.02 (ns)
Platelet distribution width	High	0.0355 (***)	0.0556 (***)	0.0222 (**)
	Low	0.0312 (***)	0.0497 (***)	0.0194 (**)
Sleeplessness/ Insomnia	Never/rarely	0.0392 (***)	0.0603 (***)	0.0251 (***)
	Sometimes	0.0331 (ns)	0.0522 (ns)	0.0206 (ns)
	Usually	0.0298 (***)	0.048 (***)	0.0184 (***)
Spells in Hospital	High	0.0424 (***)	0.0673 (***)	0.0264 (***)
	Low	0.0258 (***)	0.0406 (***)	0.0161 (***)
TS Ratio	High	0.0296 (***)	0.0473 (***)	0.0184 (***)
	Low	0.0371 (***)	0.058 (***)	0.0233 (***)
Townsend Deprivation Index	High	0.0386 (***)	0.0621 (***)	0.0231 (***)
	Low	0.0281 (***)	0.0431 (***)	0.0186 (***)
Triglycerides	High	0.0309 (***)	0.0468 (***)	0.0197 (*)
	Low	0.0358 (***)	0.0584 (***)	0.022 (*)
Urate	High	0.0336 (ns)	0.0522 (ns)	0.022 (*)
	Low	0.033 (ns)	0.0531 (ns)	0.0196 (*)
Vitamin D levels	High	0.0315 (***)	0.0474 (***)	0.0207 (ns)
	Low	0.0351 (***)	0.0578 (***)	0.0209 (ns)
Water percentage buffer (1000 m)	High	0.0305 (***)	0.0485 (***)	0.0186 (***)
	Low	0.0361 (***)	0.0568 (***)	0.0231 (***)
White blood cell leukocyte count	High	0.0343 (*)	0.0535 (ns)	0.0231 (***)
	Low	0.0324 (*)	0.0518 (ns)	0.0185 (***)

Table 3 (*continued*)

Attribute	Value	All	Carriers	Non carriers
rs429358	No alleles	0.0173 (***)	0.0147 (***)	0.0208 (ns)
	Single allele	0.0632 (***)	0.0604 (***)	
	Double Allele	0.2062 (***)	0.1976 (***)	
rs4420638	No alleles	0.0179 (***)	0.0179 (***)	0.0207 (ns)
	Single allele	0.0557 (***)	0.0611 (***)	0.0209 (ns)
	Double Allele	0.1398 (***)	0.1376 (***)	0.0766 (***)
rs6857	No alleles	0.0189 (***)	0.0229 (***)	0.0205 (**)
	Single allele	0.0562 (***)	0.0597 (***)	0.0259 (**)
	Double Allele	0.171 (***)	0.1695 (***)	0.0 (ns)
rs7412	No alleles	0.0366 (***)	0.0766 (***)	0.0208 (ns)
	Single allele	0.017 (***)	0.0162 (***)	
	Double Allele	0.0181 (**)	0.0172 (***)	
rs769449	No alleles	0.021 (***)	0.0293 (***)	0.0209 (ns)
	Single allele	0.068 (***)	0.0651 (***)	0.0 (ns)
	Double Allele	0.1857 (***)	0.1781 (***)	

Significance levels based on a chi-square proportions test of each variable unit compared to all others. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

It was hypothesized here that, in the absence of APOE4, novel factors or the same factors at a greater magnitude would be needed to meet the "AD threshold". However, nearly all other key factors identified here promoted AD more potently in APOE4 carriers. This could, in part, reflect that the case status of APOE4 carriers is more definitive in the UKB because possession of one of two copies of APOE4 lowers the age at disease onset [42]. In contrast, more non-APOE4 carriers relative to APOE4 carriers, will still go on to develop AD in later life, potentially diluting effect sizes in the non-APOE4 carrier analyses.

The highest-ranking protective factor was "Sleeplessness/Insomnia". This finding appears to contradict the idea that sleep facilitates the clearance of Aβ from the brain [43]. However, it might be consistent with studies showing that short sleep duration during midlife associates with increased Aβ deposition [44] but then in later-life, long sleep duration is associated with increased dementia risk [45]. This is further supported by a recent article that makes the distinction between sleep-initiation insomnia (increased risk) and sleep-maintenance insomnia (40% decrease) in dementia risk [46]. While UKB participants who were diagnosed with AD within two years of recruitment were excluded here, the sleep associations may also reflect the impact of AD pathology on regions such as the thalamus in the preclinical period [47].

AST:ALT ratio is commonly used in the clinic to differentiate between two main causes of chronic liver pathology; alcoholic liver disease and non-alcoholic fatty liver disease [48]. Higher AST:ALT ratios in AD and specifically correlated with CSF Aβ 1–42 levels, has been previously described [49] while ALT was inversely correlated with cognitive function. Nho *et al.* proposed that this association might reflect a global abnormality in energy metabolism including the brain or the reduced availability of glutamate, secondary to an ageing liver. A recent paper showed that the expression of human APOE4 in the liver only impaired cerebrovascular function and cognition in a mouse model [50]. It also exacerbated amyloid load when crossed with APP/PS1 mice. Both these phenotypes were rescued by plasma from young APOE3 mice. The liver is responsible for 90% of peripheral APOE and levels appear to increase with ageing [51], although APOE levels seem to be lower in E4 carriers [52]. It has been suggested that APOE4 associated changes in liver lipid metabolism, and subsequently secreted metabolites may be responsible for damage to the cerebral vasculature in AD [53]. High serum triglycerides (TG) were a risk factor here, particularly with APOE4 carriers, while our independent exploration of UKB brain MRI showed that APOE4 was associated with vascular changes such as loss of white matter integrity [54]. In the latter we also showed using the Rush Memory and Aging Project (ROSMAP) study data the degree of cerebral amyloid pathology was APOE4-dependent and immune related transcripts were more highly correlated to AD pathology levels in APOE4 carriers. This dichotomy is also consistent with a probabilistic model of AD with three variants, namely autosomal dominant, APOE4-related sporadic and APOE4 unrelated-sporadic [55].

Several variables including "Spells in Hospital", "Number of treatments/medications taken" and low ALT (contributing to higher AST:ALT ratios) collectively suggested that frailty or multimorbidity is an important component in determining who develops AD. Frailty, like dementia, can be described as a gradual loss of homeostatic mechanisms over time, and one that manifests as a multidimensional condition with physical, nutritional, psychological, and cognitive deficits [29]. Frailty and multimorbidity are related concepts [56] with most frail individuals being multimorbid but fewer multimorbid individuals meeting frailty criteria [57].

A recent study that looked at the association between a modified version of the Fried Frailty index (FFI) [58] and dementia in the UKB found that pre-frailty (one or two of the Fried Frailty criteria [58]) and frailty (three or more criteria) accounted for 9·9% and 8·6% of the 726 dementia cases (up to Jan 2017), respectively [7]. Similarly, frailty has been associated with five other chronic conditions:

multiple sclerosis, chronic fatigue syndrome; chronic obstructive pulmonary disease; connective tissue disease and diabetes using UKB data [56].

Although causal mechanisms remain unclear [59], Wallace and colleagues using the ROSMAP cohort found that frailty increased the likelihood that AD neuropathology would manifest as dementia [60].

Similarly, prominent factors such as a northernly birthplace in the UK and lower household income suggest an inverse association between socioeconomic status (SES) and AD. Poor education was similarly associated with higher risk of AD. A recent *meta*-analysis found that social class, measured by non-skilled manual occupations and poor education, but not income, was associated with the risk of dementia [61]. Persons with lower SES show faster declines in memory with ageing [62] and although not necessarily causative, social isolation, multimorbidity and low SES are common in persons living with dementia [63].

There were two indices, plasma C-reactive protein (CRP) and serum TG, where the direction of effect, low levels increasing risk, seemed counterintuitive. However recent studies do agree with these associations. Low plasma CRP at baseline was a risk in the similarly large Copenhagen General Population and the Copenhagen City Heart Studies [64]. The effect remained after adjusting for APOE4. This result is consistent with a 2008 study that CRP was lower in APOE4 carriers than controls and might reflect an influence of the locus on immune function [65].

In the case of serum TG there has been discordant studies to date, A 2020 *meta*-analysis showed no effect of serum TG levels in either AD or MCI patients [66], but a research report, showed that if TG were divided up by principal component analysis, then there was a clear association of low polyunsaturated TG with AD risk [67]. As per the findings here, this effect was greater in APOE carriers.

A weakness of the current study was the unbalanced nature of the population with only 2090 CE cases versus $\sim$ 170,000 non-AD controls. It may well be possible to use proxies for future dementia status such as hippocampal volume, neuropsychological testing performance or a family history of dementia, as demonstrated recently [68], but the current study relied on the relative surety of confirmed clinical diagnoses. The prominence of well-established risk factors such as ageing and APOE4 gives us confidence that other factors such as low serum TG and AST: ALT are important in a proportion of cases. However, ML classification models such as XGBoost do overlook the direction of causality and, as discussed above for sleep these deficits may be the earliest, preclinical effects of dementia rather than true risk factors.

5. Conclusions

This study confirms the known importance of APOE4 and ageing while suggesting that liver lipid metabolism and frailty, are important contributors to the AD. APOE4 carriers have both central and peripheral lipid anomalies that may combine to modify AD risk. The pathogenesis of AD in non-APOE4 carriers is less clear, but socioeconomic factors and multimorbidity appear relatively more important than in APOE4 carriers. Future studies using the UKB as additional AD cases occur will be importance in replicating the findings here however, these results may go some way to explaining why pathology (amyloid)-based treatments in the common sporadic cases have met with such little impact on cognitive decline. The flip side is that treatments aimed at co-morbidities and specifically liver damage may concomitantly lower AD risk.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The Authors would like to thank Dr Raymond Schwartz for his generous support of this research. The research has been conducted using the UK Biobank Resource under Application Number 49777. UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval. This approval means that researchers do not require separate ethical clearance and can operate under the RTB approval. All methods were performed in accordance with the relevant guidelines and regulations of MREC.

Funding

This work was supported through the Australian National Health & Medical Research Council (NHMRC)" to [BG]" (APP1163249) and "to [RIL]" (APP1149987), and a University of Sydney SOAR Fellowship "to [GS]". The Authors also acknowledge the generous philanthropic support of 'Eve' (a 2019 'Sydney to Hobart' yacht race entrant) supporters [GTS].

References

[1] Nichols E, Szoeke CEI, Vollset SE, Abbasi N, Abd-Allah F, Abdela J, et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019;18(1):88–106.

[2] Sutherland GT, Lim J, Srikanth V, Bruce DG. Epidemiological Approaches to Understanding the Link Between Type 2 Diabetes and Dementia. J Alzheimers Dis 2017;59(2):393–403.

[3] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020;396(10248):413–46.

[4] Graham A, Livingston G, Purnell L, Huntley J. Mild Traumatic Brain Injuries and Future Risk of Developing Alzheimer's Disease: Systematic Review and Meta-Analysis. J Alzheimers Dis 2022;87(3):969–79.

[5] Nadim R, Tang J, Dilmohamed A, Yuan S, Wu C, Bakre AT, et al. Influence of periodontal disease on risk of dementia: a systematic literature review and a meta-analysis. Eur J Epidemiol 2020;35(9):821–33.

[6] Zhang D-D, Ou Y-N, Yang L, Ma Y-H, Tan L, Feng J-F, et al. Investigating the association between cancer and dementia risk: a longitudinal cohort study. Alzheimers Res Ther 2022;14(1).

[7] Petermann-Rocha F, Lyall DM, Gray SR, Esteban-Cornejo I, Quinn TJ, Ho FK, et al. Associations between physical frailty and dementia incidence: a prospective study from UK Biobank. The Lancet Healthy Longevity 2020;1(2):e58–68.

[8] Zhang H, Greenwood DC, Risch HA, Bunce D, Hardie LJ, Cade JE. Meat consumption and risk of incident dementia: cohort study of 493,888 UK Biobank participants. Am J Clin Nutr 2021;114(1):175–84.
[9] Dimakakou E, Johnston HJ, Streftaris G, Cherrie JW. Is Environmental and Occupational Particulate Air Pollution Exposure Related to Type-2 Diabetes and Dementia? A Cross-Sectional Analysis of the UK Biobank. Int J Environ Res Public Health 2020;17(24):9581.
[10] Calvin CM, Wilkinson T, Starr JM, Sudlow C, Hagenaars SP, Harris SE, et al. Predicting incident dementia 3–8 years after brief cognitive tests in the UK Biobank prospective study of 500,000 people. Alzheimers Dement 2019;15(12):1546–57.
[11] Gong J, Harris K, Hackett M, Peters SAE, Brodaty H, Cooper M, et al. Sex differences in risk factors for cognitive decline and dementia, including death as a competing risk, in individuals with diabetes: Results from the ADVANCE trial. Diabetes Obes Metab 2021;23(8):1775–85.
[12] Deng Y-T, Kuo K, Wu B-S, Ou Y-N, Yang L, Zhang Y-R, et al. Associations of resting heart rate with incident dementia, cognition, and brain structure: a prospective cohort study of UK biobank. Alzheimers Res Ther 2022;14(1).
[13] Anaturk M, Patel R, Georgiopoulos G, Newby D, Topiwala A, de Lange AG, et al. Development and Validation of a Novel Dementia Risk Score in the UK Biobank Cohort. PsyArXiv preprints; 2021.
[14] Chen T, Guestrin C, editors. XGBoost: a scalable tree boosting system. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'16;. New York. USA: ACM Press; 2016.
[15] Lundberg S, Lee S-I, editors. A Unified Approach to Interpreting Model Predictions. NIPS2017; 2017.
[16] Moore A, Bell M. XGBoost, A Novel Explainable AI Technique, in the Prediction of Myocardial Infarction: A UK Biobank Cohort Study. Clin Med Insights Cardiol. 2022;16:11795468221133611.
[17] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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 2015;12(3):e1001779.
[18] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559–75.
[19] Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet 2021;53(9):1276–82.
[20] Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. Am J Epidemiol 2014;179(6):764–74.
[21] Radzi SFM, Karim MKA, Saripan MI, Rahman MAA, Isa INC, Ibahim MJ. Hyperparameter Tuning and Pipeline Optimization via Grid Search Method and Tree-Based AutoML in Breast Cancer Prediction. J Pers Med 2021;11(10):978.
[22] Shan G. Monte Carlo cross-validation for a study with binary outcome and limited sample size. BMC Med Inf Decis Making 2022;22(1).
[23] Ward MM. Estimating Disease Prevalence and Incidence Using Administrative Data: Some Assembly Required. J Rheumatol 2013;40(8):1241–3.
[24] McHugh ML. The Chi-square test of independence. Biochemia Medica 2013:143–9.
[25] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc: Ser B (Methodol) 1995;57(1):289–300.
[26] Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. Front Genet 2020;11:157.
[27] Soerensen M, Dato S, Tan Q, Thinggaard M, Kleindorp R, Beekman M, et al. Evidence from case-control and longitudinal studies supports associations of genetic variation in APOE, CETP, and IL6 with human longevity. Age (Dordr) 2013;35(2):487–500.
[28] Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, et al. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 1993;90(5):1977–81.
[29] Elzen PVD, Garg S, León L, Brigl M, Leadbetter EA, Gumperz JE, et al. Apolipoprotein-mediated pathways of lipid antigen presentation. Nature 2005;437(7060):906–10.
[30] Sanan DA, Weisgraber KH, Russell SJ, Mahley RW, Huang D, Saunders A, et al. Apolipoprotein E associates with beta amyloid peptide of Alzheimer's disease to form novel monofibrils. Isoform apoE4 associates more efficiently than apoE3. J Clin Invest 1994;94(2):860–9.
[31] Bennett DA, Wilson RS, Schneider JA, Evans DA, Aggarwal NT, Arnold SE, et al. Apolipoprotein E epsilon4 allele, AD pathology, and the clinical expression of Alzheimer's disease. Neurology 2003;60(2):246–52.
[32] Tachibana M, Holm ML, Liu CC, Shinohara M, Aikawa T, Oue H, et al. APOE4-mediated amyloid-beta pathology depends on its neuronal receptor LRP1. J Clin Invest 2019;129(3):1272–7.
[33] Kanekiyo T, Bu G. The low-density lipoprotein receptor-related protein 1 and amyloid-beta clearance in Alzheimer's disease. Front Aging Neurosci 2014;6:93.
[34] Ma Q, Zhao Z, Sagare AP, Wu Y, Wang M, Owens NC, et al. Blood-brain barrier-associated pericytes internalize and clear aggregated amyloid-beta42 by LRP1-dependent apolipoprotein E isoform-specific mechanism. Mol Neurodegener 2018;13(1):57.
[35] Roberts KF, Elbert DL, Kasten TP, Patterson BW, Sigurdson WC, Connors RE, et al. Amyloid-beta efflux from the central nervous system into the plasma. Ann Neurol 2014;76(6):837–44.
[36] Sagare A, Deane R, Bell RD, Johnson B, Hamm K, Pendu R, et al. Clearance of amyloid-beta by circulating lipoprotein receptors. Nat Med 2007;13(9):1029–31.
[37] Yeh FL, Wang Y, Tom I, Gonzalez LC, Sheng M. TREM2 Binds to Apolipoproteins, Including APOE and CLU/APOJ, and Thereby Facilitates Uptake of Amyloid-Beta by Microglia. Neuron 2016;91(2):328–40.
[38] Nguyen AT, Wang K, Hu G, Wang X, Miao Z, Azevedo JA, et al. APOE and TREM2 regulate amyloid-responsive microglia in Alzheimer's disease. Acta Neuropathol 2020;140(4):477–93.
[39] Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet 2017;49(9):1373–84.
[40] He Z, Guo JL, McBride JD, Narasimhan S, Kim H, Changolkar L, et al. Amyloid-beta plaques enhance Alzheimer's brain tau-seeded pathologies by facilitating neuritic plaque tau aggregation. Nat Med 2018;24(1):29–38.
[41] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261(5123):921–3.
[42] Xie L, Kang H, Xu Q, Chen MJ, Liao Y, Thiyagarajan M, et al. Sleep drives metabolite clearance from the adult brain. Science 2013;342(6156):373–7.
[43] Sindi S, Kåreholt I, Johansson L, Skoog J, Sjöberg L, Wang H-X, et al. Sleep disturbances and dementia risk: A multicenter study. Alzheimers Dement 2018;14(10):1235–42.
[44] Choe YM, Byun MS, Yi D, Lee JH, Jeon SY, Sohn BK, et al. Sleep experiences during different lifetime periods and in vivo Alzheimer pathologies. Alzheimers Res Ther 2019;11(1).
[45] Wong R, Lovier MA. Sleep Disturbances and Dementia Risk in Older Adults: Findings From 10 Years of National U.S. Prospective Data. Am J Prev Med 2023;64(6):781–7.
[46] Rub U, Stratmann K, Heinsen H, Del Turco D, Ghebremedhin E, Seidel K, et al. Hierarchical Distribution of the Tau Cytoskeletal Pathology in the Thalamus of Alzheimer's Disease Patients. J Alzheimers Dis 2016;49(4):905–15.
[47] Hall P, Cash J. What is the real function of the liver 'function' tests? Ulster Med J 2012;81(1):30–6.
[48] Nho K, Kueider-Paisley A, Ahmad S, MahmoudianDehkordi S, Arnold M, Risacher SL, et al. Association of Altered Liver Enzymes With Alzheimer Disease Diagnosis, Cognition, Neuroimaging Measures, and Cerebrospinal Fluid Biomarkers. JAMA Netw Open 2019;2(7):e197978.
[49] Liu C-C, Zhao J, Fu Y, Inoue Y, Ren Y, Chen Y, et al. Peripheral apoE4 enhances Alzheimer's pathology and impairs cognition by compromising cerebrovascular function. Nat Neurosci 2022;25(8):1020–33.
[50] Gee JR, Ding Q, Keller JN. Modulation of apolipoprotein E and interleukin-1beta in the aging liver. Exp Gerontol 2005;40(5):409–15.
[51] Simon R, Girod M, Fonbonne C, Salvador A, Clément Y, Lantéri P, et al. Total ApoE and ApoE4 isoform assays in an Alzheimer's disease case-control study by targeted mass spectrometry (n=669): a pilot assay for methionine-containing proteotypic peptides. Mol Cell Proteomics 2012;11(11):1389–403.

[52] Giannisis A, Patra K, Edlund AK, Nieto LA, Benedicto-Gras J, Moussaud S, et al. Brain integrity is altered by hepatic APOE epsilon4 in humanized-liver mice. Mol Psychiatry 2022;27(8):3533–43.
[53] McCorkindale AN, Mundell HD, Guennewig B, Sutherland GT. Vascular Dysfunction Is Central to Alzheimer's Disease Pathogenesis in APOE e4 Carriers. Int J Mol Sci 2022;23(13):7106.
[54] Frisoni GB, Altomare D, Thal DR, Ribaldi F, van der Kant R, Ossenkoppele R, et al. The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. Nat Rev Neurosci 2022;23(1):53–66.
[55] Hanlon P, Nicholl BI, Jani BD, Lee D, McQueenie R, Mair FS. Frailty and pre-frailty in middle-aged and older adults and its association with multimorbidity and mortality: a prospective analysis of 493 737 UK Biobank participants. Lancet Public Health 2018;3(7):e323–32.
[56] Vetrano DL, Palmer K, Marengoni A, Marzetti E, Lattanzio F, Roller-Wirnsberger R, et al. Frailty and Multimorbidity: A Systematic Review and Meta-analysis. J Gerontol A Biol Sci Med Sci. 2019;74(5):659-66.
[57] Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56(3):M146-56.
[58] Chu NM, Xue QL, McAdams-DeMarco MA, Carlson MC, Bandeen-Roche K, Gross AL. Frailty-a risk factor of global and domain-specific cognitive decline among a nationally representative sample of community-dwelling older adult U.S. Medicare beneficiaries. Age Ageing 2021;50(5):1569–77.
[59] Wallace LMK, Theou O, Godin J, Andrew MK, Bennett DA, Rockwood K. Investigation of frailty as a moderator of the relationship between neuropathology and dementia in Alzheimer's disease: a cross-sectional analysis of data from the Rush Memory and Aging Project. Lancet Neurol 2019;18(2):177–84.
[60] Bodryzlova Y, Kim A, Michaud X, Andre C, Belanger E, Moullec G. Social class and the risk of dementia: A systematic review and meta-analysis of the prospective longitudinal studies. Scand J Public Health 2022;14034948221110019.
[61] Avila-Rieger J, Turney IC, Vonk JMJ, Esie P, Seblova D, Weir VR, et al. Socioeconomic Status, Biological Aging, and Memory in a Diverse National Sample of Older US Men and Women. Neurology 2022;99(19):E2114–24.
[62] Sadarangani T, Perissinotto C, Boafo J, Zhong J, Yu G. Multimorbidity patterns in adult day health center clients with dementia: a latent class analysis. BMC Geriatr 2022;22(1):514.
[63] Hegazy SH, Thomassen JQ, Rasmussen IJ, Nordestgaard BG, Tybjaerg-Hansen A, Frikke-Schmidt R. C-reactive protein levels and risk of dementia-Observational and genetic studies of 111,242 individuals from the general population. Alzheimers Dement 2022;18(11):2262–71.
[64] Haan MN, Aiello AE, West NA, Jagust WJ. C-reactive protein and rate of dementia in carriers and non carriers of Apolipoprotein APOE4 genotype. Neurobiol Aging 2008;29(12):1774–82.
[65] Liu Y, Zhong X, Shen J, Jiao L, Tong J, Zhao W, et al. Elevated serum TC and LDL-C levels in Alzheimer's disease and mild cognitive impairment: A meta-analysis study. Brain Res 2020;1727:146554.
[66] Bernath MM, Bhattacharyya S, Nho K, Barupal DK, Fiehn O, Baillie R, et al. Serum triglycerides in Alzheimer disease: Relation to neuroimaging and CSF biomarkers. Neurology. 2020;94(20):e2088-e98.
[67] Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet 2019;51(3):404–13.