Abstract

Background

Considerable uncertainty remains regarding associations of multiple risk factors with Alzheimer's disease (AD). We aimed to systematically screen and validate a wide range of potential risk factors for AD.

Methods and Materials

Among 502,493 participants from the UK Biobank (UKB), baseline data were extracted for 4171 factors spanning ten different categories. Phenome-wide association analyses (PheWAS) and time-to-event analyses were conducted to identify factors associated with both polygenic risk scores for AD (AD-PRS) and AD diagnosis at follow-up. We performed Two-sample Mendelian randomization (MR) analysis to further assess their potential causal relationships with AD and imaging association analysis to discover underlying mechanisms.

Results

We identified 39 factors significantly associated with both AD-PRS and risk of incident AD, where higher levels of education, body size, basal metabolic rate, fat-free mass, computer use, and cognitive
functions were associated with a decreased risk of developing AD, and food selective intake and more outdoor exposures with an increased risk of developing AD. The identified factors were also associated with AD-related brain structures including hippocampus, entorhinal cortex and inferior/middle temporal cortex, and 21 of them were further supported by MR evidence.

Conclusions

This study, for the first time, comprehensively and rigorously assessed the effects of wide-ranging risk factors on AD. Strong evidence was found for body fat-free mass, basal metabolic rate, computer use, food selective intake, and outdoor exposures as new risk factors for AD. Integration of genetic, clinical, and neuroimaging information may help prioritize risk factors and prevention targets for AD.

Keywords

Alzheimer’s disease, Phenome-Wide Association Study, Mendelian Randomization Analysis, Time-to-event analysis, UK Biobank

Introduction

Alzheimer’s disease (AD) is the leading contributor to the global burden of disease, due to its high prevalence and disabling consequences(1). Many risk factors have been identified to be associated with AD. Recent publications are focusing on validating limited sets of risk factors, and data from large-scale cohorts are increasingly used to acquire knowledge on actionable strategies that target those factors and prevent the disease onset (2, 3). However, these studies have been mainly carried out with hypothesis-driven designs. Additional factors may remain overlooked or unknown.

AD is a hereditary disease. The heritability is estimated to be between 60% and 80%(4). Recent genome-wide association studies (GWAS) conducted by Schwartzentruber et al. (5) and Bellenguez et al.(6) have respectively identified genetic variants spanning 37 and 75 risk-loci. By using summary statistics from these large-scale GWAS, multiple factors were discovered to share genetic architecture with AD and possibly involved in the pathology of the disease(7, 8, 9, 10). This revealed that genetic association is an effective tool for identifying AD risk factors and identifying those that may have causal effects on AD.

A phenome-wide association study (PheWAS) is a type of hypothesis-free analysis where a broad range of phenotypes can be examined in genetic association with a disease. Through applying a polygenic risk score (PRS) of AD (AD-PRS) as a proxy for AD risk, the associations of a wide array of both established and undiscovered non-genetic factors with AD can be systematically screened (11). This approach has received relatively little attention owing to a lack of resources with sufficient variety and volume. Large-scale datasets such as the UK Biobank (UKB) now provide an unparalleled opportunity for this approach (12). Such efforts have a significant advantage compared to traditional observational studies that require long-term follow-up and large sample sizes, as statistical power can be dramatically increased by exploiting genetic and phenotypic information. Other analytic techniques can be subsequently performed to rigorously assess the putative associations and reduce false-positive findings, including traditional longitudinal analysis, and Mendelian randomization (MR)
In the present study, using phenotypic and genomic data from 502,493 UKB participants, we first generated AD-PRS and conducted a PheWAS to investigate the effect and significance of associations between AD-PRS and all the available risk factors. A total of 84 factors were found to be associated with AD-PRS and followed by time-to-event analyses to evaluate their associations with clinically significant AD at follow-up. Next, two-sample MR analysis was performed to evaluate potential causal relationships between the identified factors and AD. Finally, we conducted imaging analysis to explore the underlying biological mechanisms for the factors associated with both AD-PRS and incident AD.

Methods and materials

Study population

The UKB is a population-based cohort of more than 500,000 participants in the United Kingdom between 2006 and 2010 (14). The UKB has research tissue bank approval from the North West Multi-centre Research Ethics Committee (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics) and provided oversight for this study. Participation is voluntary, and participants are free to withdraw at any time without giving any reason. Written informed consent was obtained from all participants. Genetic and phenotypic data were obtained at baseline. Clinical outcomes including AD diagnoses were available over a follow-up period from 2007 to 2020 via hospital in-patient records, death certificates, primary care records, and self-reports. The initial sample included 502,493 participants aged between 38 and 73. Data acquisition and analyses in the present study were conducted under UKB Application #19542. AD-PRS generation, PheWAS, and MR analysis only included populations of European ancestry to reduce the impact of population structures on genetic data analysis. To ensure adequate power, all participants, regardless of ethnicity, were included in time-to-event and imaging analyses as many as possible.

AD-PRS generation

Genotype data were available for all 502,493 participants in UKB. Detailed genotyping and quality control procedures by the UKB are available in a previous publication (15). We excluded single-nucleotide polymorphisms (SNPs) with call rates <95%, minor allele frequency <0.1%, deviation from the Hardy–Weinberg equilibrium with p<1×10^-10, and selected subjects that were estimated to have recent British ancestry based on self-report information and principal components analyses of the genotypes and have no more than ten putative third-degree relatives in the kinship table. After the quality control procedures, we obtained a total of 591,050 SNPs and 337,199 participants.

We calculated the PRS with the summary statistics from a meta-analysis of GWAS from four large AD consortia(16), which included a total of 7,055,881 SNPs and 54,062 individuals (17,008 cases and 37,054 controls).

PRSiCe was the software (www.PRSice.info) for calculation of AD-PRS. We used P-value-informed clumping with a cutoff of r^2 = 0.1 in a 250-kb window in the analysis (17). P-thresholds for scoring were set at p<0.0005, p<0.001, p<0.005, p<0.01, p<0.05, p<0.1, p<0.5 and p<1 (18).
The factors entering PheWAS consisted of 10 broad categories (containing 4171 variables), which were: 1) Sociodemographic; 2) Physical measures; 3) Lifestyle and environment; 4) Health conditions; 5) Mental health; 6) Medications and operations; 7) Cognitive function; 8) Sex-specific factors; 9) Employment; 10) Early life factors. These variables were from four categories (Population characteristics, Assessment Centre, Online follow-up, and Health-related outcomes) in the UKB showcase and were re-categorized to a small extent based on the framework of the showcase. For further details see, Figure 1 and Supplementary Data Table 1-2.

![Figure 1](https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf)

**Figure 1.** Risk factors’ overview. There were 4171 factors of 10 categories included. The top panel demonstrates the number of factors with median and range of sample size in each category with the size of pie chart section indicating the ratio of each category. The down panel is the bar plot illustrating the effect sizes of all included factors. The y axis represents the mean standard effect size (absolute value) across AD-PRS generated with all eight p thresholds. Colors indicate categories.

### Imaging data of brain structures

Quality-controlled T1-weighted magnetic resonance imaging (MRI) data were used for studying associations between factors and brain structures. Details of the process procedure can be found in the UKB protocol ([https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf](https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf)) and **Supplementary Methods**. Surface templates were utilized to extract imaging derived phenotypes (IDP) referring to atlas regions’ surface volume ([19](#)) and FreeSurfer’s aseg was used to extract subcortical regions ([20](#)). FreeSurfer aparc (ID=192) and ASEG (ID=190) atlas corresponding to 68 cortical regions and 41 subcortical regions were applied in the current study.

### Statistical analysis

**PheWAS**

The PHESANT package in R was used to test the PheWAS associations. The PHESANT’s automated rule-based method was described in detail in a previous publication ([21](#)). In brief, decision rules were based on the variable type, and each variable was categorized as one of four data types: continuous, ordered categorical, unordered categorical or binary. Normality of continuous data was ensured by an
inverse normal rank transformation prior to testing. In the current study, AD-PRSs were set as independent variables, and selected factors were set as dependent variables, with age, gender, genotyping array, the first 10 genetic principal components, and the assessment center as covariates included in the model. Overall, 4171 factors × 8 AD-PRS = 33368 tests across factors and AD-PRS p thresholds were corrected altogether by Benjamini-Hochberg procedure (false discovery rate (FDR)-correction) (q < 0.05). Only factors found significantly associated with AD-PRS at a minimum of four PRS variant p thresholds would be evaluated subsequently. For direct comparison of the results between linear and logistic regression models, standardized regression coefficients were estimated as effect sizes (β) for both types of models and were reported with log-transformed odds ratio for binary dependent variables. Two-sided statistical tests were applied in all analyses.

Time-to-event analysis

Multivariable Cox proportional hazard regression models were used to examine the association of the PheWAS-selected factors with incident AD. Follow-up was calculated in person-years from date of recruitment until date of first incident AD diagnosis, death, loss to follow-up, or the last date of hospital admission data available, whichever came first. The model was adjusted for age, gender, and APOE4 carrier. The proportional hazards assumption was evaluated using tests of Schoenfeld residuals. In the main analysis, participants with prevalent dementia diagnoses, including AD diagnosis at baseline or before age 50, were excluded, and were limited to those with follow-up periods of ≥3 years.

Mendelian randomization analysis

‘TwosampleMR’ package in R was used to conduct two-sample MR analyses. GWAS summary data for factors were acquired from the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/) and summary data from a recent large AD GWAS meta-analysis (7) was used as outcome dataset. The inverse variance weighted (IVW) method was the primary method for conducting MR. MR-PRESSO (pleiotropy residual sum and outlier) was mainly used for detecting potential pleiotropy and correcting IVW estimates. See reverse MR and more details in Supplementary Methods.

Associations between factors and brain structures

Linear regression models were applied to investigate the association of selected factors with brain morphometric measures. Covariates were age, gender, APOE4 carrier, and imaging scanning site. FDR corrections were conducted for multiple comparisons among cortical and subcortical regions respectively. Hippocampus (22, 23), entorhinal (24, 25), inferior temporal (26, 27, 28), and middle temporal (26) cortex out of FreeSurfer aparc and ASEG atlas were selected as AD-related brain structures based on previous publications, which were also supported with UKB data (see Supplementary Results).

Data availability

The data used in our study is from UKB with restrictions applied. Data are used under license and thus are not publicly available. Access to UKB data can be requested through a standard protocol (https://www.ukbiobank.ac.uk/register-apply/). The summary statistics of AD GWAS for PRS calculation is available at http://www.pasteur-lille.fr/en/recherche/u744/Igap_stage1.zip. GWAS summary statistics used in MR are detailed in Supplementary Information.
Code availability

Scripts used to perform the analyses are available at https://github.com/HaloForest/UKB_PWAS.

Results

A summary of the 4171 factors that entered the analytic pipeline is shown in Figure 1. The analytic pipeline is illustrated in Supplementary Figure 1 with participants inclusion in each analysis step shown in Supplementary Figure 2. An overview of the results from the four analytic steps is shown in Figure 2.

Figure 2. Factors associated with AD-PRS across analytic steps. Main results (39 out of 84 factors) were shown in bold. Green and red cells indicate decreased and increased risk of AD, while blank cells indicate no results supporting associations between factors and AD. Initially, 84 factors survived PheWAS; then, 51 survived time-to-event analyses; then, 21 of 51 factors were further supported by MR and 39 factors showed significant results in imaging analysis. Except for non-significant associations discovered between factors and AD, other conditions shown by blank cells includes: a Factors are duplicates thus not entering subsequent analysis steps; b Factors have no non-incident AD thus not entering time-to-event analysis and subsequent steps; c Factors shows contradictory effect directions between time-to-event analysis and PheWAS; d Factor shows no associations with brain structures in imaging analysis; e Factors show associations with inconsistent effect directions between imaging previous analyses. Details can be found in Supplementary Data Table 1,4,6,9-12.

Abbreviations: AD=Alzheimer’s Disease, BMI=Body Mass Index, FI=Fluid Intelligence, IPAQ=international physical activity questionnaire, MR=Mendelian Randomization, MET=Metabolic Equivalent of Energy, NM=Numeric Memory, TM=Trail Making, PM=Pair Matching.
PheWAS and time-to-event analyses identified 51 factors spanning six categories to be consistently associated with AD.

Initially, 84 factors survived PheWAS. This included 8 sociodemographics, 33 physical measures, 18 lifestyle and environment, 5 health conditions, 8 medications and operations, 11 cognitive functions, and 1 sex-specific factor (standardized coefficients $\beta$: -0.0758–0.369, $p_{\text{FDR}}$: 5.60×10$^{-5}$–0.050, Figure 3, Supplementary Data Table 1). All the significant associations showed an identical effect direction for each of 84 factors. Population baseline characteristics of PheWAS can be found in Supplementary Table 1. The complete PheWAS results for 4171 factors are presented in Supplementary Data Table 2. In total, 22.4% of cognitive phenotypes and 11.7% of physical measures showed significant results in PheWAS. The proportions were lower for phenotypes of sociodemographic (8.9%), health conditions (5.95%), sex-specific factors (3.22%), lifestyle and environment (1.01%), and medications and operations (0.45%), and no phenotypes of employment, mental health, or early life factors showed significant results. After excluding 3 duplicate factors with smaller sample sizes (Body mass index (BMI), weight, and fluid intelligence score) and 2 factors where all incident AD were from the disease group (vascular dementia and unspecified dementia), 79 factors entered time-to-event analysis.

Figure 3. Heatmap for the factors significantly associated with AD-PRS. The shown factors were significantly associated with AD-PRS at a minimum of four p thresholds for AD-PRS. Shades of cells indicate the standardized effect sizes ($\beta$) between each AD-PRS and each phenotype with a darker color denoting a larger effect size. Cells with an asterisk indicated significant associations after FDR correction. Factors marked with * means duplication. Detail can be found in Supplementary Data.
Of these, 51 factors survived time-to-event analyses, while 26 showed non-significant associations with incident AD, and 2 had inconsistent effect directions between PheWAS and time-to-event analyses. Baseline descriptions are presented in Supplementary Data Table 3. The complete results of time-to-event analysis for the 79 factors are provided in Supplementary Data Table 4. We herein reported 22 factors of them with details in main text (Table 1).

<table>
<thead>
<tr>
<th>Factors</th>
<th>Sample size</th>
<th>HR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualifications: College/University</td>
<td>105450/244615</td>
<td>0.740</td>
<td>(0.659,0.831)</td>
<td>3.81E-07</td>
</tr>
<tr>
<td>Qualifications: none</td>
<td>65688/284377</td>
<td>1.371</td>
<td>(1.248,1.507)</td>
<td>5.80E-11</td>
</tr>
<tr>
<td>Age completed full time education</td>
<td>243608</td>
<td>0.963</td>
<td>(0.939,0.988)</td>
<td>4.46E-03</td>
</tr>
<tr>
<td><strong>Physical measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>352351</td>
<td>0.982</td>
<td>(0.972–0.992)</td>
<td>6.02E-04</td>
</tr>
<tr>
<td>Weight</td>
<td>352487</td>
<td>0.989</td>
<td>(0.985,0.992)</td>
<td>5.59E-10</td>
</tr>
<tr>
<td>Standing height</td>
<td>352737</td>
<td>0.977</td>
<td>(0.970,0.983)</td>
<td>4.06E-11</td>
</tr>
<tr>
<td>Hip circumference</td>
<td>352873</td>
<td>0.988</td>
<td>(0.983,0.993)</td>
<td>1.04E-05</td>
</tr>
<tr>
<td>Basal metabolic rate</td>
<td>347141</td>
<td>0.9998</td>
<td>(0.9997,0.9999)</td>
<td>1.60E-11</td>
</tr>
<tr>
<td>Whole body fat-free mass</td>
<td>347123</td>
<td>0.974</td>
<td>(0.966,0.981)</td>
<td>1.73E-11</td>
</tr>
<tr>
<td><strong>Lifestyle and environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No selective intake of eggs/dairy/wheat/sugar</td>
<td>269915/82376</td>
<td>0.672</td>
<td>(0.613,0.738)</td>
<td>3.89E-17</td>
</tr>
<tr>
<td>Never eat sugar or foods containing sugar</td>
<td>61792/269915</td>
<td>1.353</td>
<td>(1.218,1.502)</td>
<td>1.48E-08</td>
</tr>
<tr>
<td>Time spent using computer</td>
<td>276808</td>
<td>0.904</td>
<td>(0.865,0.945)</td>
<td>6.66E-06</td>
</tr>
<tr>
<td>Time spent outdoors in summer</td>
<td>320857</td>
<td>1.022</td>
<td>(1.001,1.043)</td>
<td>3.59E-02</td>
</tr>
<tr>
<td>Time spent outdoors in winter</td>
<td>279866</td>
<td>1.052</td>
<td>(1.026,1.079)</td>
<td>8.92E-05</td>
</tr>
<tr>
<td><strong>Health conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorders of lipoprotein metabolism</td>
<td>88272/265227</td>
<td>1.173</td>
<td>(1.071,1.285)</td>
<td>6.18E-04</td>
</tr>
<tr>
<td>Delirium</td>
<td>3268/350238</td>
<td>13.82</td>
<td>(12.43–15.34)</td>
<td>4.06E-11</td>
</tr>
</tbody>
</table>
Qualifications of college/university degree was associated with both lower AD-PRS (β: -0.0366 to -0.0027, P<10^{-6}) and decreased risk of incident AD (hazard ratio [HR]: 0.74; 95% confidence interval [95%CI]: 0.659–0.831; P = 3.81×10^{-10}). In contrast, no qualification attainment was associated with higher AD-PRS (β: 0.0183–0.0392, P<10^{-10}) and increased risk of incident AD (1.371(1.248–1.507), P = 5.80×10^{-10}). Consistent with these results, older age at completing full-time education was associated with lower AD-PRS (β: -0.0270 to -0.0241, P<10^{-10}) and decreased risk of AD (0.963(0.939–0.989), P = 4.46×10^{-10}).

Multiple associations were identified for physical measures. Higher levels of global body measures were associated with lower AD-PRS and decreased risk of incident AD, including BMI (β: -0.0103 to -0.0072, P<10^{-8}), weight (β: -0.0091 to -0.0052, P<10^{-8}), standing height (β: -0.0099 to -0.0043, P<10^{-8}), hip circumference (β: -0.0086 to -0.0058, P<10^{-8}), basal metabolic rate (β: -0.0064 to -0.0061, P<10^{-8}), and whole body fat-free mass (β: -0.0057 to -0.0038, P<10^{-8}). Significant associations were also found for multiple physical measures in different parts of the body (Figure 2, Figure 3).

No selective intake of four types of food (eggs, dairy, wheat, and sugar) was associated with lower AD-PRS and lower risk of developing AD (β: -0.0294 to -0.0246, P<10^{-8}), while never eating sugar or foods/drink containing sugar was associated with higher AD-PRS and increased risk of AD (β: 0.0151–0.0175, P<10^{-8}). Interestingly, more time spent using a computer was associated with...
both lower AD-PRS and decreased risk of incident AD ($\beta$: -0.0251 to -0.0128, $P_{FDR}$: 5.36x10^{-9} to 2.77x10^{-2}), and more time spent outdoors in summer/winter were found to be associated with higher AD-PRS and elevated risk of AD ($\beta$: 0.0129-0.0244, $P_{FDR}$: 2.88x10^{-9} - 4.80x10^{-2}; 1.022-1.052 (1.001–1.079); $P = 3.59x10^{-8}$ - 8.92x10^{-10}).

Associations with both higher AD-PRS and increased risk of AD were also identified for certain health conditions and medications including disorders of lipoprotein metabolism and other lipidaemias ($\beta$: 0.0099-0.0688, $P_{FDR}$: 4.31x10^{-40} - 2.43x10^{-5};1.173(1.071–1.285), $P = 6.18x10^{-14}$), delirium ($\beta$: 0.042-0.1988, $P_{FDR}$: 6.12x10^{-17}- 2.43x10^{-3}; 13.81(12.43–15.34), $P = 4.06x10^{-11}$) and simvastatin intake ($\beta$: 0.0170-0.0566, $P_{FDR}$: 3.27x10^{-16} – 6.26x10^{-2}; 1.15(1.03–1.28), $P = 1.18x10^{-12}$).

Better performance in different cognitive function tests were consistently associated with both lower AD-PRS and decreased risk of AD, including higher total scores in 'Fluid intelligence' test ($\beta$: -0.0508 to -0.0125, $P_{FDR}$: 4.75x10^{-11}- 2.05x10^{-3}; 0.798(0.761–0.836), $P = 1.99x10^{-21}$), more numbers of digits memorized in the 'Numeric memory' test ($\beta$: -0.0301 to -0.0073, $P_{FDR}$: 1.26x10^{-2} – 4.13x10^{-2}; 0.770(0.719–0.826), $P = 2.59x10^{-13}$), more numbers of symbol digit matches made correctly in the 'Symbol digit substitution' test ($\beta$: -0.0213 to -0.0053, $P_{FDR}$: 1.56x10^{-4} – 1.63x10^{-2}; 0.834(CI: 0.811–0.858), $P = 1.44x10^{-15}$), fewer numbers of incorrect matches per round in the 'Pairs matching' test ($\beta$: 0.0024-0.0098, $P_{FDR}$: 1.14x10^{-2} – 3.95x10^{-2}; 1.084(1.067–1.100), $P = 2.42x10^{-24}$), and shorter interval in alphanumeric path (trail #2) in the 'Trail Making' test ($\beta$: 0.0073-0.02553, $P_{FDR}$: 2.78x10^{-6} – 3.07x10^{-3}; 1.016(1.013–1.019), $P = 1.25x10^{-25}$). Other cognitive factors also showed consistent effects (Figure 2, Figure 3).

Mendelian randomization evidence for associations between 21 factors and AD

The potentially causal effects on AD were found for 21 of 51 factors which showed the same effect direction as those of PhewAS and time-to-event analyses. These contained the aforementioned factors including 2 factors in education (College or University degree [OR$_{IVW}$: 0.475, $P_{IVW}$: 2.01x10^{-5}] and no qualifications attained [OR$_{IVW}$: 4.00, $P_{IVW}$: 1.51x10^{-4}]), 4 in physical measures (Weight, standing height, basal metabolic rate, and whole body fat-free mass[OR$_{IVW}$: 0.767-0.881, $P_{IVW}$: 3.41x10^{-3}-0.044]), 1 in lifestyle and environment (Time spent using computer[OR$_{IVW}$: 0.713, $P_{IVW}$: 0.041]), and 2 in cognitive function (Fluid intelligence score [OR$_{MR-PRESSO corrected}$ : 0.916, $P_{MR-PRESSO corrected}$: 0.040] and interval in alphanumeric path (trail #2) in the ‘Trail Making test’ [OR$_{IVW}$: 1.726, $P_{IVW}$: 7.65x10^{-3}] ) (see major results in Figure 4). These associations were all significant at $P<0.05$ (IVW or correction by the MR-PRESSO method) without directional pleiotropy and the estimates were consistent across MR methods. Full results can be found in Supplementary Data Table 6-8.
Figure 4. Mendelian randomization estimates of factors in relation to AD risk. Major factors (14 of 21) found associated with AD in MR analysis are presented here. Full results can be found in Supplementary Data Table 6 and details can be found in Supplementary Figure 6-42. Abbreviations: AD=Alzheimer's Disease, CI=Confidence Interval, FI=Fluid Intelligence, IVW=Inverse-Variance Weighted, MR=Mendelian Randomisation, MR-PRESSO=MR-pleiotropy residual sum and outlier TM=Trail Making, OR=Odds Ratio
Associations between 39 identified factors and AD neuroimaging hallmarks

Out of 51 factors, 39 showed significant associations with volumes of multiple AD-related brain structures including the hippocampus, the entorhinal, inferior temporal, and middle temporal cortex (Supplementary Figure 3) with the consistent effect directions as those of PheWAS and time-to-event analysis. These included all the 21 factors with MR evidence and other 18 factors such as no selective intake of four types of food (eggs, dairy, wheat, and sugar), never eating sugar or foods/drink containing sugar, time spent outdoors in summer/winter, disorders of lipoprotein metabolism, simvastatin intake, and fluid intelligence score. We plotted the significant associations for those among factors with largest effects (Figure 5). Full lists of statistic reports (β and P_{FDR}) for associations between 51 factors and 68 cortical/41 subcortical regions is presented in Supplementary Data Table 9-12.

Figure 5. Significant associations between factors and AD-related regions. Thirty-nine out of 51 factors showed consistent associations between cortical and subcortical regions, of which ten with largest effect were presented. AD-related regions were marked on the left hemisphere. The color bars indicate the effect size of factors on brain regions. Associations between 51 factors and AD-related regions are shown in the heatmap in Supplementary Figure 3. Statistic details can be found in Supplementary Data Table 9-12.

Discussion

Integrating genetic, clinical, and neuroimaging information from the UKB, this study for the first time comprehensively evaluated the associations between a wide array of risk factors and AD. The most robust findings were found for 21 associations where higher education attainment, greater body
size, greater fat-free mass, faster basal metabolic rate, more computer use, and better cognitive status were associated with decreased risk of developing AD. Strong evidence was also found for other 18 factors where selective food intake and more time spent outdoors were associated with increased risk of developing AD. Collectively, fat-free mass, basal metabolic rate, computer use, selective food intake, and outdoor exposures were identified as novel risk factors for AD.

The findings on education and cognitive function were in line with common views of risk factors for AD. The long-standing evidence from the literature has validated that higher education and better cognitive function are associated with lower risks of developing AD(29, 30). Recent research using PRS also found increased education associated with decreased odds of AD diagnosis (31) and cognitive status greatly improved prediction of AD in 3-8 years(32). These results emphasize their important roles in the risk factor profiles of AD.

Fat-free mass and basal metabolic rate were new risk factors for AD. The direct evidence on the associations between two factors and AD was limited. One case-control study observed that loss of lean mass was associated with brain atrophy and cognitive performance(33) and a recent cohort study found that decreased lean mass was an indicator of increased all-cause dementia risk in older adults(34). Little study on AD involved “basal metabolic rate”, partially because the variable was not commonly measured in research. In our study, significant associations identified in PheWAS indicated that they shared strong genetic background with AD and time-to-event analysis confirmed their clinical association with disease’s occurrence. We also ran sensitivity analyses and the significant results remained after adjusting for covariates (Supplementary Data Table 15,17). Evidence from MR further strengthened their associations with AD and imaging association analyses suggested that they influenced the risk of AD by altering AD-specific brain areas. Future research should investigate whether preventing fat-free mass loss and maintaining the basal metabolic rate in older adults reduces AD risk.

Our study also adds to the body of evidence on the link between computer use and AD, which is lacking in the literature. We thought the association was likely driven by the association between education and AD. However, the significant results remained after we adjusted for covariates such as education level (Supplementary Data Table 15,17). One study previously observed less computer use in cognitively impaired individuals compared to the cognitively unimpaired (35) and less computer use was related to smaller hippocampal volumes and worse cognitive performance (35, 36), which partly supported our findings. One explanation is that computer use may act as a form of cognitive training to maintain cognitive function and prevent the onset of AD [38]. Because our findings were consistent across all four analytic steps, further investigations are warranted to assess its role in AD prevention as a potentially modifiable risk factor.

Novel findings on associations for other lifestyles were also supported by strong evidence from our research. Significant results from PheWAS, time-to-event, and imaging analysis consistently showed that people who had no selective intake of eggs, dairy, wheat, sugar, or products containing them had lower risks of AD. In contrast, people who never ate sugar had higher risks of AD. Both results supplement the current dietary research on AD (2, 37) that a balanced diet cannot be neglected. Three analytic steps also consistently found that time spent outdoors was associated with increased risk of AD. This was indirectly supported by a recent work also conducted by our team where time spent outdoors was associated with elevated risks of developing all-cause dementia after adjusting for a series of covariates(38). We noted that associations between these factors and AD were not bolstered by MR. As estimates of MR represent lifelong average effects of genetic variants, we think that MR
may not be interpreted in the same way as those from a relatively briefer life period and that absence of MR support does not refute the potential importance of the factors (39). Considering that those factors are not previously identified and they have mild-to-moderate effects, it is worthwhile to thoroughly investigate their roles in AD in separate studies.

One may find the factors’ effects on brain structures were relatively small in our imaging analysis. This was consistent with recent findings on brain-wide association studies that the effects of brain–phenotype associations were smaller than previously thought. It was estimated that the top 1% largest of all possible brain-wide associations only reached a standard correlations (bivariate linear $|r|$) value greater than 0.06(40). Thus, small effect sizes did not imply null findings and factors may be associated with AD via mechanisms other than altering the morphology of brain structures.

In PheWAS, we noticed differential association between multiple categories of risk factors. More than 10% of cognitive phenotypes and physical measures were significantly associated with AD, likely reflecting their overlapping genetic architecture with AD. In contrast, phenotypes of other categories such as lifestyle and environment and medications and operations showed much lower proportions of significance, suggesting that AD shares its genetic architecture with only a small proportion of them. One potential reason for this discrepancy is a relative lack of signal in GWASes of behavioral variables like lifestyle, which contrasts with endophenotypes like cognitive and physical conditions. When putting phenotypes of various categories together, the results for endophenotypes will most likely come out on top.

A novel concept provides new insights into PheWAS. Desikan RS et al.(41) recently developed and validated a novel polygenic risk score, the polygenic hazard score (PHS), which was found to predict the onset age of AD dementia. A higher PHS was linked to faster clinical decline, faster atrophy of AD-specific brain regions, and more amyloid/tau deposition (42). As a result, we believe this score represents the genetic architecture of AD progression, rather than the PRS, which only represents the risk of AD. Future research could use PHS to conduct a phenome-wide scan and identify factors that share genetic backgrounds with AD progression. This may reveal different factor profiles from the current study, providing more clues on disease-modifying therapies and effective intervention.

Several limitations should be considered when interpreting our findings. First, our phenome-wide study was restricted by the available variables from UKB database. Variables such as air pollution and blood glucose measures were not included, although they had been previously found to be associated with AD (43, 44, 45, 46). Second, because PheWAS looks for phenotypes that are associated with genetic architecture of AD, our study is less likely to uncover AD risk factors that have no or a weak genetic link to the disease. This limitation was highlighted by the current PheWAS's stringent filtering criteria. They may obscure certain associations that are potentially noteworthy if studied individually. In our PheWAS, for example, sleep duration, daytime naps, and time spent watching television were all significantly associated with AD-PRS in at least two P thresholds, but due to conservative thresholds, they were excluded from further analysis. Thus, more nuanced analyses are required. Third, although we performed additional analysis adjusting for covariates such as education and BMI, we were unable to triangulate the relationships between all the identified factors. Future studies may seek to distinguish between these factors’ independent and mediation effects to decipher potential pathways that underpin the associations.

In conclusion, leveraging phenotypic and genomic data from over 500,000 individuals in UKB, we used a novel four-step approach to systematically screen and rigorously assess associations of a broad
array of risk factors with AD and found strong evidence for body fat-free mass, basal metabolic rate, computer use, food selective intake, and outdoor exposures as new risk factors for AD. Integration of genetic, clinical, and neuroimaging information may help prioritize risk factors and prevention targets for AD.

Uncited reference
29, 30, 40.

Acknowledgments and disclosures
Data acquisition and analyses were conducted using the UK Biobank Resource under approved project #19542. Participation of the UK Biobank participants is gratefully appreciated. We also thank the UK Biobank team for collecting and preparing data for analyses. This work was made possible by the generous sharing of GWAS summary statistics. We thank the IEU GWAS database project as well as the individuals and institutes joining this project for offering the summary statistics of the GWAS data. We also thank the Alzheimer Disease Genetics Consortium (ADGC), the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), the European Alzheimer's Disease Initiative (EADI), and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERAD), the Alzheimer's Disease Sequencing Project (ADSP), the Alzheimer's disease working group of the Psychiatric Genomics Consortium (PGC-ALZ), and the International Genomics of Alzheimer's Project (IGAP) for providing summary statistics for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control participants, the patients, and their families. The i–Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728. The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations.

JTY had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: JTY, WC, QD; Acquisition, analysis, or interpretation of data: All authors; Drafting of the manuscript: SDC, WZ, YZL; Critical revision of the...
manuscript for important intellectual content: All authors; Statistical analysis: SDC, WZ, YZL, WC; Obtained funding: JTY, WC, JFF; Administrative, technical, or material support: WC, JFF, QD, JTY; Supervision: JFF, WC, JTY, QD. All authors read and approved the final manuscript.

J.T. Yu serves as Editor-in-Chief for Brain Disorders, Associate Editor-in-Chief for Annals of Translational Medicine, and Senior Editor for Journal of Alzheimer Disease. All other authors report no biomedical financial interests or potential conflicts of interest.

This study is based on publicly available data with different levels of accessibility. The UK Biobank and individual studies within each genome-wide association study received approval from an appropriate institutional review board, and informed consent was obtained from participants or from a caregiver, legal guardian, or other proxy.

This study was supported by grants from the Science and Technology Innovation 2030 Major Projects (2022ZD0211600), National Natural Science Foundation of China (82071201, 81971032), Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), Research Start-up Fund of Huashan Hospital (2022QD002), Excellence 2025 Talent Cultivation Program at Fudan University (3030277001), and ZHANGJIANG LAB, Tianqiao and Chrissy Chen Institute, and the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of Ministry of Education, Fudan University. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Supplementary Material

Download all supplementary files

- Download Acrobat PDF file (5MB)
- Download spreadsheet (13MB)
- Download spreadsheet (20KB)

Recommended articles

References

1. G.B.D.D. Collaborators
   Lancet Neurol, 18 (2019), pp. 88-106
   View Record in Scopus    Google Scholar


T.G. Richardson, S. Harrison, G. Hemani, G. Davey Smith
An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome
Elife, 8 (2019)

Polygenic dissection of major depression clinical heterogeneity
Mol Psychiatry, 21 (2016), pp. 516-522

D.A. Lawlor, K. Tilling, G. Davey Smith
Triangulation in aetiological epidemiology
Int J Epidemiol, 45 (2016), pp. 1866-1886

Multimodal population brain imaging in the UK Biobank prospective epidemiological study
Nat Neurosci, 19 (2016), pp. 1523-1536

The UK Biobank resource with deep phenotyping and genomic data

Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease
Nat Genet, 45 (2013), pp. 1452-1458

Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores


An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest
Neuroimage, 31 (2006), pp. 968-980

Whole brain segmentation: automated labeling of neuroanatomical structures in the human
brain

Software Application Profile: PHESANT: a tool for performing automated phenome scans in
UK Biobank
Int J Epidemiol, 47 (2018), pp. 29-35

22 G. Halliday
Pathology and hippocampal atrophy in Alzheimer's disease
Lancet Neurol, 16 (2017), pp. 862-864

23 Y. Mu, F.H. Gage
Adult hippocampal neurogenesis and its role in Alzheimer's disease
Mol Neurodegener, 6 (2011), p. 85

MRI of entorhinal cortex in mild Alzheimer's disease

Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease
J Neurosci, 16 (1996), pp. 4491-4500

Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented
elderly predict decline to Alzheimer's disease

Inferior temporal tau is associated with accelerated prospective cortical thinning in clinically
normal older adults
Neuroimage, 220 (2020), Article 116991

Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer's
disease
J Alzheimers Dis, 24 (2011), pp. 547-557

29 S. Norton, F.E. Matthews, D.E. Barnes, K. Yaffe, C. Brayne
Potential for primary prevention of Alzheimer's disease: an analysis of population-based data
Lancet Neurol, 13 (2014), pp. 788-794

Y. Stern
Cognitive reserve in ageing and Alzheimer's disease
Lancet Neurol, 11 (2012), pp. 1006-1012


Predicting incident dementia 3-8 years after brief cognitive tests in the UK Biobank prospective study of 500,000 people

J.M. Burns, D.K. Johnson, A. Watts, R.H. Swerdlow, W.M. Brooks
Reduced lean mass in early Alzheimer disease and its association with brain atrophy
Arch Neurol, 67 (2010), pp. 428-433

C. Cui, R.H. Mackey, C.E. Shaaban, L.H. Kuller, O.L. Lopez, A. Sekikawa
Associations of body composition with incident dementia in older adults: Cardiovascular Health Study-Cognition Study
Alzheimers Dement, 16 (2020), pp. 1402-1411

Less Daily Computer Use is Related to Smaller Hippocampal Volumes in Cognitively Intact Elderly
J Alzheimers Dis, 52 (2016), pp. 713-717

Passively-Measured Routine Home Computer Activity and Application Use Can Detect Mild Cognitive Impairment and Correlate with Important Cognitive Functions in Older Adulthood
J Alzheimers Dis, 81 (2021), pp. 1053-1064

Dietary patterns during adulthood and cognitive performance in midlife: The CARDIA study
Neurology, 92 (2019), pp. e1589-e1599

Time spent in outdoor light is associated with the risk of dementia: a prospective cohort study of 362094 participants
BMC medicine, 20 (2022), p. 132

An Exposure-Wide and Mendelian Randomization Approach to Identifying Modifiable Factors for the Prevention of Depression

Reproducible brain-wide association studies require thousands of individuals
Nature, 603 (2022), pp. 654-660

Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score
PLoS medicine, 14 (2017), Article e1002258

L. Calderon-Garciduenas, R. Villarreal-Rios
Living close to heavy traffic roads, air pollution, and dementia
Lancet, 389 (2017), pp. 675-677

X. Zhang, X. Chen, X. Zhang
The impact of exposure to air pollution on cognitive performance

A.M. de Matos, M.P. de Macedo, A.P. Rauter
Bridging Type 2 Diabetes and Alzheimer's Disease: Assembling the Puzzle Pieces in the Quest for the Molecules With Therapeutic and Preventive Potential