nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01585-x

# Identifying modifiable factors and their joint effect on dementia risk in the UK Biobank

Received: 16 December 2022

Accepted: 10 March 2023

Published online: 06 April 2023

Check for updates

Yi Zhang[1,8], Shi-Dong Chen[1,8], Yue-Ting Deng[1], Jia You[2,3,4], Xiao-Yu He[1], Xin-Rui Wu[1], Bang-Sheng Wu[1], Liu Yang[1], Ya-Ru Zhang[1], Kevin Kuo[1], Jian-Feng Feng[2,3,4,5], Wei Cheng[1,2,3,4], John Suckling[6], A. David Smith[7] & Jin-Tai Yu[1]

Previous hypothesis-driven research has identified many risk factors linked to dementia. However, the multiplicity and co-occurrence of risk factors have been underestimated. Here we analysed data of 344,324 participants from the UK Biobank with 15 yr of follow-up data for 210 modifiable risk factors. We first conducted an exposure-wide association study and then combined factors associated with dementia to generate composite scores for different domains. We then evaluated their joint associations with dementia in a multivariate Cox model. We estimated the potential impact of eliminating the unfavourable profiles of risk domains on dementia using population attributable fraction. The associations varied by domain, with lifestyle (16.6%), medical history (14.0%) and socioeconomic status (13.5%) contributing to the majority of dementia cases. Overall, we estimated that up to 47.0%–72.6% of dementia cases could be prevented.

Dementia is a leading cause of disability and dependency in older people[1]. Given the lack of effective treatments, exploring modifiable risk factors to design preventive measures is an important, perhaps only current way available to reduce dementia burden[2]. In the past decades, hypothesis-driven methods have been commonly adopted to explore the modifiable factors[3–12] that could be integrated into dementia prevention packages. However, these methods have several limitations. First, single-exposure analyses are highly likely to produce overestimated effect sizes and type I errors due to the interconnected nature of risk factors[13]. Second, these studies contain selective reporting constraining reproducibility[14]. Third, investigating one or a handful of risk factors at a time cannot reflect the synergistic effects of exposures, which is also important owing to the multifactorial nature of late-life dementia[4]. Lastly, single-exposure analyses have not shed light on the overall contribution of risk factors to dementia, which has currently gained considerable attention as recent randomized controlled trials have shown the effectiveness of multidomain lifestyle interventions for dementia prevention[2].

An exposure-wide association study (EWAS) is a hypothesis-free strategy that systematically and agnostically investigates the relationship between multiple variables and a single outcome. Researchers have successfully applied this technique to complex diseases other than dementia, including depression[15], HIV[16] and diabetes[17]. By investigating a wide range of exposures simultaneously, EWAS validates established factors from previous studies with reduced bias and false-positive findings[13,18] while enabling the discovery of novel risk factors[19]. The analytic rational, similar to genome-wide association study (GWAS), employs standardized analytical procedures and generates results with greater robustness compared with hypothesis-driven approaches[18]. Moreover, by constructing composite scores[4] and calculating a population

[1]Department of Neurology and Institute of Neurology, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, National Center for Neurological Disorders, Shanghai, China. [2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [3]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Fudan University, Ministry of Education, Shanghai, China. [4]Fudan ISTBI—ZJNU Algorithm Centre for Brain-Inspired Intelligence, Zhejiang Normal University, Jinhua, China. [5]MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. [6]Department of Psychiatry, University of Cambridge, Cambridge, UK. [7]Department of Pharmacology, University of Oxford, Oxford, UK. [8]These authors contributed equally: Yi Zhang, Shi-Dong Chen. e-mail: jintai_yu@fudan.edu.cn

https://doi.org/10.1038/s41562-023-01585-x

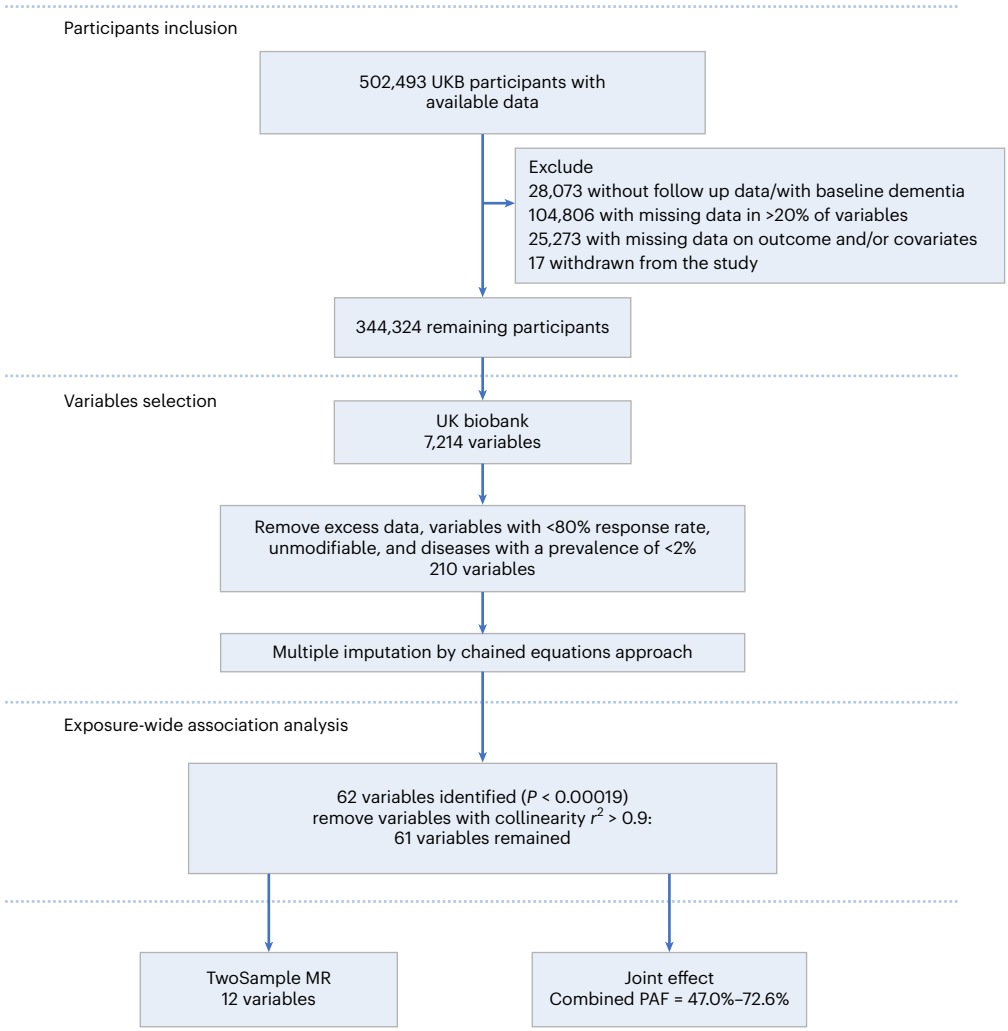



**Fig. 1 | Overview of analytic design.** Analytical procedure to identify modifiable risk factors associated with incident dementia in the UK Biobank. TwoSample MR, 2-sample Mendelian randomization.

attributable fraction (PAF)[20], the joint effects of multiple risk factors can be composed and their contributions to dementia prevalence can be determined.

In this study, leveraging phenotypic and genomic data from over 300,000 UK Biobank (UKB) participants, we first conducted an EWAS in the field of dementia, aiming to comprehensively identify risk factors. Then, by combining risk factors to create composite scores for different domains, we investigated the joint effects of multidomain factors on dementia. Finally, we quantified the PAFs for each domain and in total for dementia to uncover the power of preventative approaches (Fig. 1).

## Results

### Identification in prospective exposure-wide analysis

Of the 344,324 individuals included in this study, 54.2% were female. The mean (s.d.) age was 56.01 (8.04) yr. A total of 4,654 participants developed dementia and the mean (s.d.) age at diagnoses was 72.8 (5.9) yr (see Supplementary Table 5 and Supplementary Note for population characteristics). In total, 62 out of 210 risk factors were significantly associated with dementia (Supplementary Table 7 and Fig. 2). Among them, 46 factors showed potentially detrimental effects and 16 were protective. Among the top 10 factors, 3 reduced liability for dementia: hand grip strength ($Z = -13.74$, hazard ratio (HR) = 0.49, 95% confidence interval (CI) = 0.44–0.54, $P = 5.59 \times 10^{-43}$), lung function ($Z = -12.11$, HR = 0.56, 95% CI = 0.51–0.61, $P = 9.42 \times 10^{-34}$) and average total household income before tax (31,000 to 51,999: $Z = -11.46$, HR = 0.62, 95% CI = 0.57–0.67, $P = 2.04 \times 10^{-30}$; 52,000 to 100,000: $Z = -12.03$, HR = 0.50, 95% CI = 0.45–0.56, $P = 2.60 \times 10^{-33}$); and 7 had increased liability for dementia: disability ($Z = 19.77$, HR = 2.35, 95% CI = 2.16–2.56, $P = 5.79 \times 10^{-87}$), stroke ($Z = 15.10$, HR = 2.50, 95% CI = 2.22–2.82, $P = 1.57 \times 10^{-51}$), diabetes ($Z = 14.20$, HR = 1.86, 95% CI = 1.70–2.02, $P = 9.53 \times 10^{-46}$), depression ($Z = 12.96$, HR = 1.77, 95% CI = 1.63–1.93, $P = 2.01 \times 10^{-38}$), traumatic brain injury ($Z = 11.55$, HR = 2.94, 95% CI = 2.45–3.53, $P = 7.46 \times 10^{-31}$), unemployed ($Z = 10.03$, HR = 1.96, 95% CI = 1.72–2.24, $P = 1.08 \times 10^{-23}$) and coronary heart disease (CHD, $Z = 9.85$, HR = 1.51, 95% CI = 1.39–1.63, $P = 6.96 \times 10^{-23}$).

None of the early life factors was significant after correction (Figs. 2a and 3). Figure 3, Supplementary Figs. 1–6 and Extended Data Fig. 1 report the EWAS results in subgroups. Similar pattern of associations was observed when results were stratified by age, sex, apolipoprotein E (*APOE*) ε4 status and follow-up time. 'Natural environment percentage, buffer 1,000 m' was excluded from the investigation of collinearity.

Mendelian randomization (MR) supported 12 factors (Supplementary Tables 11 and 12, and Extended Data Fig. 2), which included traumatic subdural haemorrhage (odds ratio (OR) = 1.06, $P = 0.037$), two factors of media use (time spent watching TV (OR = 1.38, $P = 0.035$) and time spent using a computer (OR = 0.70, $P = 0.032$)), five factors of education (College or University degree, A levels/AS levels or equivalent,

https://doi.org/10.1038/s41562-023-01585-x

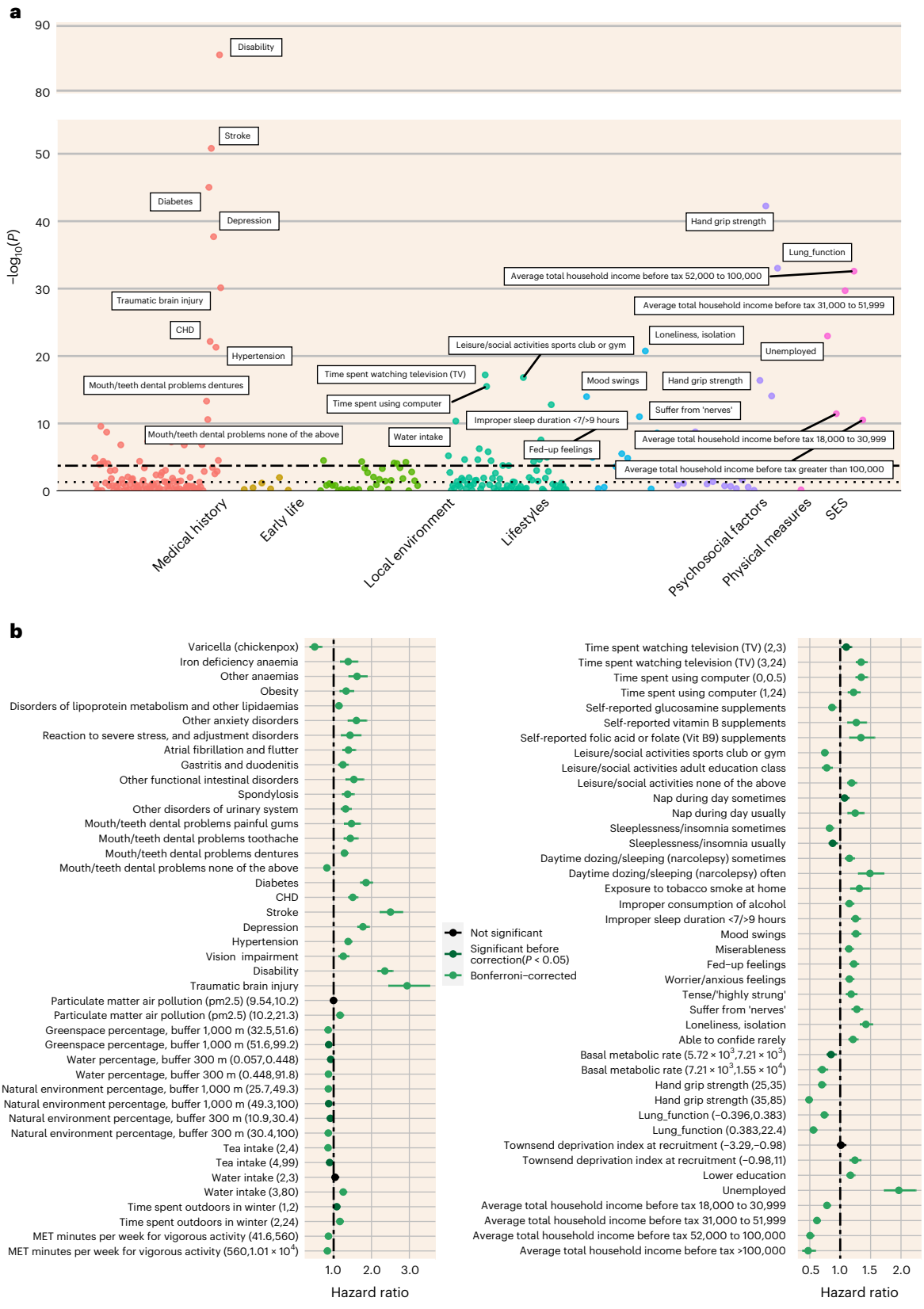



**Fig. 2 | Associations between modifiable risk factors and incident dementia.** **a**, The *x* axis shows the category domains and the *y* axis represents statistical significance (that is, $-\log_{10}$ of the *P* value). The horizontal dotted line indicates the significance threshold of $P < 0.05$, while the dashed line indicates threshold after correcting for multiple testing (Bonferroni correction, $P < 1.87 \times 10^{-4}$). *Z*-tests were used to assess statistical significance and derive *Z* statistics and corresponding two-sided *P* values ($N = 344{,}324$). A set of top risk factors were annotated. The full results are available in Supplementary Table 7 and Figs. 1–6, and Fig. 3. **b**, Dots represent hazard ratios, horizontal lines indicate corresponding 95% CIs. Hazard ratios were calculated using Cox proportional hazards regression analysis after adjustments for age at baseline age, sex, *APOE* ε4 status and assessment centre ($N = 344{,}324$).

https://doi.org/10.1038/s41562-023-01585-x

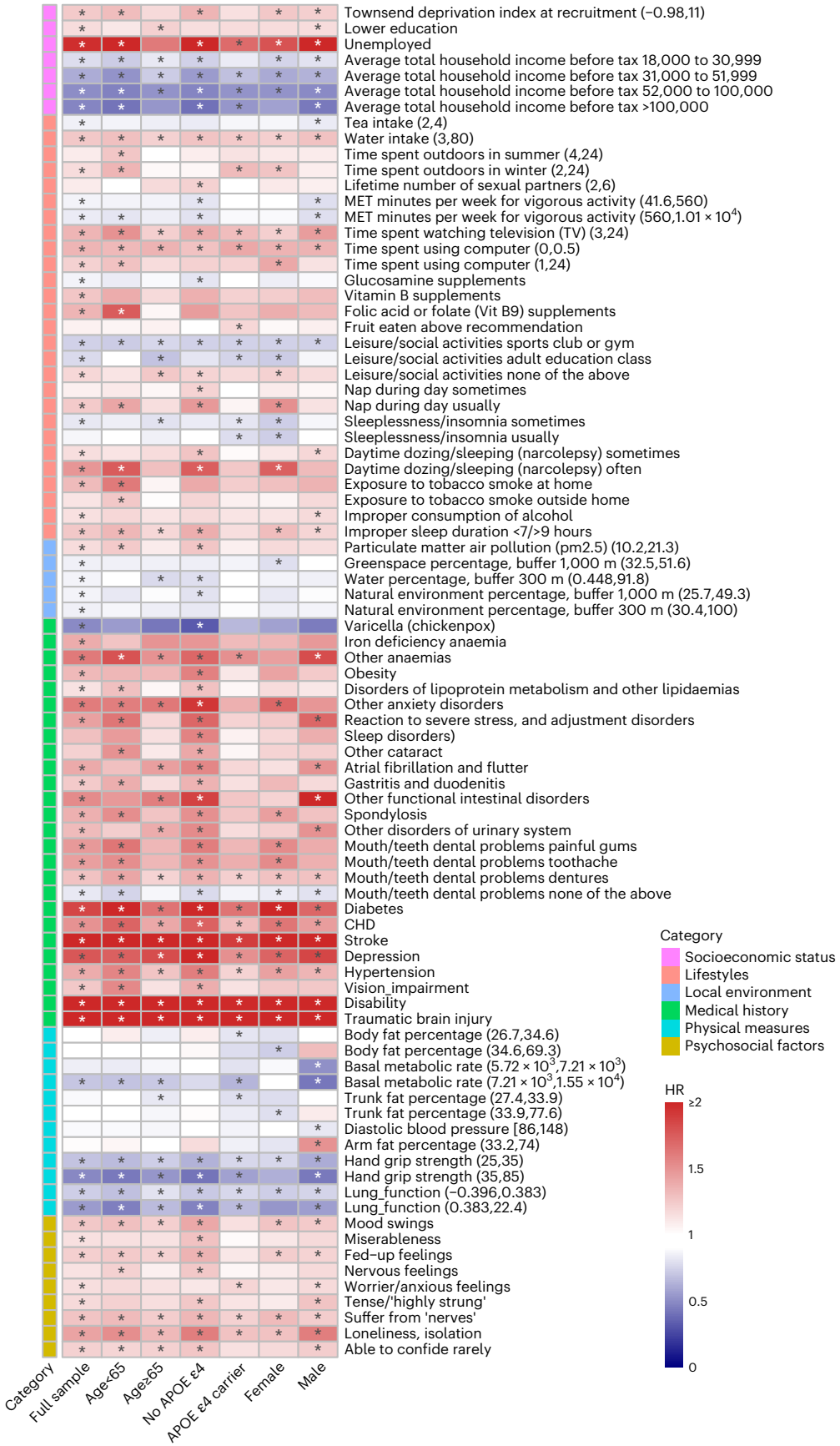



**Fig. 3 | Summary heat map for significant factors in EWAS analysis across the full sample and subgroups.** Models were adjusted for baseline age, sex, *APOE* ε4 status and assessment centre. The colour of cells indicates the effect sizes (HR) between each risk factor and incident dementia ($N = 344{,}324$). Asterisks in cells represent significant associations after correction for multiple testing (Bonferroni-corrected, $P < 1.87 \times 10^{-4}$).

https://doi.org/10.1038/s41562-023-01585-x

| | Total no. of participants | No. of cases of dementia | HR (95% CI) | *P* value | *P* for trend |
|---|---|---|---|---|---|
| Lifestyles | | | | | |
| Favourable | | | 1 (reference) | | |
| Intermediate | 97,597 | 1,211 | 1.29 (1.18–1.40) | <0.001 | |
| Unfavourable | 100,467 | 1,842 | 1.62 (1.49–1.76) | <0.001 | <0.001 |
| Medical history | | | | | |
| Favourable | | | 1 (reference) | | |
| Intermediate | 49,045 | 725 | 1.09 (0.99–1.19) | 0.08 | |
| Unfavourable | 98,228 | 2,008 | 1.52 (1.41–1.64) | <0.001 | <0.001 |
| Local environment | | | | | |
| Favourable | | | 1 (reference) | | |
| Intermediate | 100,349 | 1,389 | 1.13 (1.05–1.23) | 0.002 | |
| Unfavourable | 96,807 | 1,354 | 1.16 (1.07–1.26) | <0.001 | <0.001 |
| Psychosocial factors | | | | | |
| Favourable | | | 1 (reference) | | |
| Intermediate | 93,492 | 1,250 | 1.13 (1.04–1.22) | 0.003 | |
| Unfavourable | 100,308 | 1,392 | 1.18 (1.09–1.28) | <0.001 | <0.001 |
| Physical measures | | | | | |
| Favourable | | | 1 (reference) | | |
| Intermediate | 95,088 | 1,194 | 1.32 (1.21–1.44) | <0.001 | |
| Unfavourable | 97,859 | 1,615 | 1.85 (1.62–2.10) | <0.001 | <0.001 |
| SES | | | | | |
| Favourable | | | 1 (reference) | | |
| Intermediate | 96,575 | 1,321 | 1.13 (1.04–1.24) | 0.006 | |
| Unfavourable | 90,051 | 1,773 | 1.29 (1.18–1.41) | <0.001 | <0.001 |

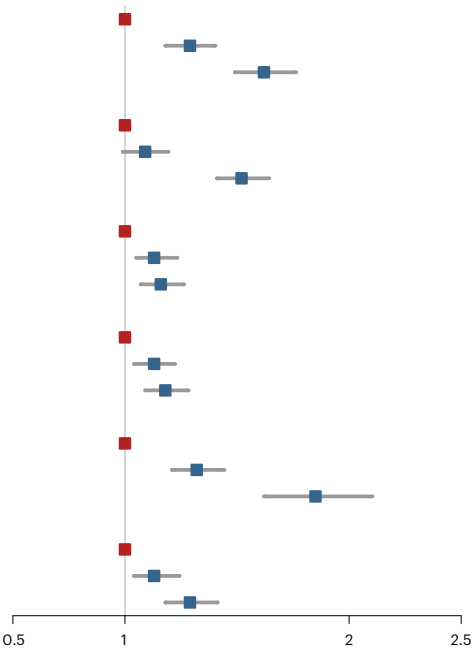



**Fig. 4 | Associations between six domains and dementia.** The favourable profile was set as reference in each domain. The associations were estimated by applying Cox model including all six domains after mutual adjustment and adjustment for age, sex, *APOE* ε4 status and assessment centre. Dots represent hazard ratios; horizontal lines indicate corresponding 95% CIs. *Z*-tests were used to assess statistical significance and derive *Z* statistics and corresponding two-sided *P* values.

O levels/GCSEs or equivalent, adult education class (OR = 0.07–0.47, $P = 1.03 \times 10^{-5}$–0.032) and no qualification (OR = 4.10, $P = 9.20 \times 10^{-5}$)), two factors of physical measures (basal metabolic rate and FEV1 (OR = 0.75–0.84, $P = 1.55 \times 10^{-5}$–0.047)), moderate physical activity (OR = 1.96, $P = 0.013$) and household income (OR = 0.75, $P = 0.037$). Supplementary Table 11 reports the inverse-variance weighted results after outlier removal. Results for other methods are shown in Supplementary Table 11. No evidence of heterogeneity was observed, and MR Egger suggested no horizontal pleiotropy.

### Joint effects of identified factors on dementia

Compared with the favourable profile, intermediate and unfavourable profiles of lifestyles (respective $Z = 5.56, 11.27$, HR = 1.29,1.62, 95% CI = 1.18–1.40, 1.49–1.76, $P = 2.69 \times 10^{-8}$, $<2 \times 10^{-16}$), local environment ($Z = 3.13, 3.46$, HR = 1.13, 1.16, 95% CI = 1.05 − 1.23, 1.07 − 1.26, $P = 0.002, 0.0006$), psychosocial factors ($Z = 3.00, 4.16$, HR = 1.13, 1.18, 95% CI = 1.04 − 1.22, 1.09 − 1.28, $P = 0.003, 3.25 \times 10^{-5}$), physical measures ($Z = 6.12, 9.16$, HR = 1.32, 1.85, 95% CI = 1.21 − 1.44, 1.62 − 2.10, $P = 9.10 \times 10^{-10}$, $<2 \times 10^{-16}$), socioeconomic status (SES) ($Z = 2.76, 5.64$, HR = 1.13, 1.29, 95% CI = 1.04 − 1.24, 1.18 − 1.41, $P = 0.006, 1.71 \times 10^{-8}$) and an unfavourable profile of medical history ($Z = 10.86$, HR = 1.52, 95% CI = 1.41 − 1.64, $P < 2 \times 10^{-16}$) significantly increased the risk of dementia. A trend toward significance was found for all six domains (*P* for trend < 0.001) (Fig. 4). The pattern of results, except for psychosocial factors, was nearly identical in the sensitivity analysis (Supplementary Table 13).

Moreover, we found statistically significant interactions between the weighted scores for the six domains and genetic risks based on *APOE* genotypes (*P* for interaction $<2 \times 10^{-6}$). Further analyses stratified by genetic risk showed that favourable profiles of the six domains were generally related to decreased dementia risk across genetic categories (Table 1). Among participants with the highest genetic risk (ε2ε4, ε3ε4 or ε4ε4), a favourable profile of lifestyles (HR = 0.62, 95% CI = 0.55–0.71, $P = 7.16 \times 10^{-13}$), medical history (HR = 0.79, 95% CI = 0.70–0.89, $P = 5.42 \times 10^{-5}$), physical measures (HR = 0.62, 95% CI = 0.51–0.77, $P = 8.69 \times 10^{-6}$), SES (HR = 0.77, 95% CI = 0.67–0.88, $P = 1.76 \times 10^{-4}$) and local environment (HR = 0.88, 95% CI = 0.77–0.99, $P = 0.044$) significantly decreased the risk of dementia.

### PAF estimates for the six domains in dementia prevention

When shifting all unfavourable profiles to intermediate and favourable ones (Model 1), PAF estimation suggests that 47.0% of dementia cases might be prevented, increasing to 72.6% when shifting all factors to the favourable tertile (Model 2; Table 2, and Supplementary Tables 14 and 15). In a more conservative situation (Model 1), the greatest impact in terms of prevention was estimated to come from medical history, leading to a 12.3% reduction in incidence of dementia. Other domains were responsible for 10% (lifestyles), 8.4% (SES), 8.2% (physical measures), 4.9% (psychosocial factors) and 3.2% (local environment) of dementia cases. If more thorough elimination was possible, lifestyles contributed to the most dementia cases (16.6%), followed by medical history (14%), SES (13.5%), physical measures (12.8%), psychosocial factors (9.0%) and local environment (6.7%). When selecting factors using a machine learning method (hierarchical clustering) instead of correlation analysis after the EWAS, no significant change was found for our findings (Supplementary Tables 16 and 17 and Fig. 7, and Extended Data Fig. 3).

## Discussion

Consistent with the multifactorial aetiology of dementia, we first identified a diverse array of risk factors from six domains: lifestyles, medical history, SES, physical measures, psychosocial factors and local environment. Poorer profiles in these six domains independently increased the risk of developing dementia over 15 yr of follow-up. Significant interactions were found between genetic risk and the six domains, and favourable lifestyles, medical history, physical measures, SES

https://doi.org/10.1038/s41562-023-01585-x

**Table 1 | Risk of incident dementia according to categories of six domains within each genetic risk category**

| | Low genetic risk (ε2ε2 or ε2ε3) | | | Intermediate genetic risk (ε3ε3) | | | High genetic risk (ε2ε4, ε3ε4 or ε4ε4) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Lifestyles** | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable |
| (*N*/*n*) | 10,586/48 | 10,409/66 | 10,734/109 | 49,068/293 | 48,647/400 | 49,690/645 | 21,679/352 | 21,527/522 | 22,050/746 |
| HR (95% CI) | 0.63 (0.44–0.89) | 0.79 (0.58–1.07) | 1.00 | 0.61 (0.53–0.70) | 0.75 (0.66–0.85) | 1.00 | 0.62 (0.55–0.71) | 0.83 (0.74–0.92) | 1.00 |
| *P* value | 0.010 | 0.129 | Reference | $1.14\times10^{-11}$ | $8.46\times10^{-6}$ | Reference | $7.16\times10^{-13}$ | 0.0009 | Reference |
| **Medical history** | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable |
| (*N*/*n*) | 15,956/48 | 5,311/39 | 10,462/136 | 74,661/367 | 24,078/232 | 48,666/739 | 33,117/538 | 10,759/330 | 21,380/752 |
| HR (95% CI) | 0.39 (0.28–0.56) | 0.58 (0.40–0.83) | 1.00 | 0.57 (0.50–0.64) | 0.66 (0.57–0.76) | 1.00 | 0.79 (0.70–0.89) | 0.82 (0.72–0.94) | 1.00 |
| *P* value | $9.91\times10^{-8}$ | 0.003 | Reference | $<2.0\times10^{-16}$ | $3.14\times10^{-8}$ | Reference | $5.42\times10^{-5}$ | 0.004 | Reference |
| **Local environment** | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable |
| (*N*/*n*) | 10,622/70 | 10,665/77 | 10,442/76 | 49,241/376 | 50,023/478 | 48,141/484 | 21,731/480 | 22,257/584 | 21,268/556 |
| HR (95% CI) | 0.95 (0.67–1.33) | 1.02 (0.73–1.41) | 1.00 | 0.79 (0.69–0.91) | 0.96 (0.84–1.09) | 1.00 | 0.88 (0.77–0.99) | 0.99 (0.88–1.12) | 1.00 |
| *P* value | 0.762 | 0.922 | Reference | 0.001 | 0.530 | Reference | 0.044 | 0.935 | Reference |
| **Psychosocial factors** | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable |
| (*N*/*n*) | 10,960/65 | 9,923/72 | 10,846/86 | 51,122/401 | 46,555/433 | 49,728/504 | 22,396/552 | 20,710/526 | 22,150/542 |
| HR (95% CI) | 0.80 (0.57–1.11) | 0.87 (0.63–1.19) | 1.00 | 0.78 (0.68–0.89) | 0.93 (0.82–1.06) | 1.00 | 0.90 (0.80–1.02) | 1.03 (0.91–1.17) | 1.00 |
| *P* value | 0.175 | 0.387 | Reference | 0.0002 | 0.294 | Reference | 0.093 | 0.612 | Reference |
| **Physical measures** | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable |
| (*N*/*n*) | 11,073/70 | 10,226/69 | 10,430/84 | 50,891/390 | 46,999/436 | 49,515/512 | 22,846/436 | 20,943/437 | 21,467/747 |
| HR (95% CI) | 0.43 (0.25–0.73) | 0.57 (0.36–0.91) | 1.00 | 0.55 (0.44–0.69) | 0.77 (0.64–0.92) | 1.00 | 0.62 (0.51–0.77) | 0.72 (0.61–0.85) | 1.00 |
| *P* value | 0.002 | 0.017 | Reference | $1.21\times10^{-7}$ | 0.004 | Reference | $8.69\times10^{-6}$ | $7.30\times10^{-5}$ | Reference |
| **SES** | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable | Favourable | Intermediate | Unfavourable |
| (*N*/*n*) | 11,853/44 | 10,278/79 | 9,598/100 | 54,807/287 | 48,202/437 | 44,396/614 | 24,313/328 | 21,327/575 | 19,616/717 |
| HR (95% CI) | 0.78 (0.54–1.15) | 0.98 (0.72–1.33) | 1.00 | 0.80 (0.68–0.92) | 0.85 (0.75–0.96) | 1.00 | 0.77 (0.67–0.88) | 0.93 (0.83–1.04) | 1.00 |
| *P* value | 0.209 | 0.892 | Reference | 0.003 | 0.010 | Reference | $1.76\times10^{-4}$ | 0.212 | Reference |

Hazard ratios were calculated using Cox proportional hazards regression analysis after mutual adjustment and adjustment for sex, age and assessment centre. *N*, number of individuals at risk; *n*, number of dementia cases. *Z*-tests were used to assess statistical significance and derive *Z* statistics and corresponding two-sided *P* values.

and local environment were associated with a lower risk of dementia among people with the highest genetic risk (ε2ε4, ε3ε4 or ε4ε4). Overall, 47.0%–72.6% of dementia cases in the present population could be prevented by adhering to more favourable profiles in these six domains, with lifestyles (16.6%), medical history (14%) and SES (13.5%) accounting for most dementia cases.

EWAS identified 62 correlated factors. In line with the literature, modifiable factors such as health conditions involving the cardiovascular system[21,22], metabolic syndrome[23], neuropsychiatry[24], gastrointestinal system[25], mouth/teeth[9] and traumatic brain injury;[8,26] media use and sleep duration;[27] physical activity;[27] hand grip strength;[10] lung function;[6,28] and education[29] were among the top correlates. Relatively unexplored factors including second-hand smoke[7,30], varicella[12], spondylosis, disability, time spent outdoors in winter[31], emotions like miserableness and tense, unemployment, glucosamine and able to confide were also identified.

When only eliminating the unfavourable profiles, medical history was the largest PAF contributor to dementia (12.3%). A more thorough elimination of risk factors showed that adhering to favourable lifestyles had the greatest potential, with a 16.6% reduction in dementia incidence. Similarly, in The Lancet commission's life course prevention model, diseases accounted for half of the included risk factors[26]. Other studies found a monotonic relationship between increasing multimorbidity and dementia risk[32], with one study suggesting that the association between cardiometabolic multimorbidity and dementia risk is greater than that of lifestyle factors combined[33]. Our results suggest that public health programmes should focus primarily on other associated illnesses if lifestyle interventions cannot be fully implemented. In addition to commonly mentioned diabetes, cardiovascular diseases and neuropsychiatric disorders, attention should also be paid to people with mouth and teeth problems, disability and gastrointestinal issues.

Adhering to favourable lifestyles was estimated to reduce dementia by 16.6%. Similarly, the FINGER trial showed a beneficial effect of a 2 yr multidomain lifestyle intervention on cognition[2]. Lifestyle factors targeted by previous randomized controlled trials typically included dietary, exercise and cardiovascular health modifications, but in our analysis, consumption of vegetables, fruit, other recommended diets

**Table 2 | Weighted and unweighted PAF for the six domains**

| | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|
| **Domains** | **Unweighted PAF** | **Communality** | **Weighted PAF** | **Unweighted PAF** | **Communality** | **Weighted PAF** |
| Lifestyles | 0.185 | 0.180 | 0.100 | 0.323 | 0.175 | 0.166 |
| Medical history | 0.226 | 0.399 | 0.123 | 0.273 | 0.224 | 0.140 |
| Local environment | 0.059 | 0.483 | 0.032 | 0.130 | 0.040 | 0.067 |
| Psychosocial factors | 0.089 | 0.416 | 0.049 | 0.175 | 0.104 | 0.090 |
| Physical measures | 0.151 | 0.166 | 0.082 | 0.249 | 0.128 | 0.128 |
| SES | 0.154 | 0.356 | 0.084 | 0.263 | 0.329 | 0.135 |
| Overall weighted PAF | | | 0.470 | | | 0.726 |

Weighted PAF was calculated after considering the overlap between risk factors. In Model 1, we shifted the unfavourable profiles to intermediate and favourable ones. In model 2, we shifted all factors to the favourable tertile.

and a healthy diet score were not significant in the EWAS and were therefore not included in the lifestyle score construction. This is consistent with the reported conflicting effects of diet[34]. Our lifestyle score contained less-discussed factors, such as time spent outdoors in winter, time of media use, mineral supplementation and sleep duration. These factors require additional confirmation before moving towards clinical trials. Of course, for all factors, adherence to interventions or changes to lifestyles is necessary to gain and then sustain the benefits they confer. Further work here is clearly important[2].

Low SES also contributed greatly to dementia (PAF: 13.5%). Previous studies found that SES has a greater impact than comorbidities and lifestyle factors on disparities in dementia incidence among black and white older people[35]. Due to the higher rate of missed diagnoses in low SES areas, our calculations of the associated PAF may be underestimates[36]. Socioeconomic factors like education can affect brain growth and maturation. Lower SES has been linked to altered brain networks and decreased cortical thickness in middle-aged adults, making them more susceptible to dementia and other cognitive impairments[37]. Although it is clear that low SES is associated with a higher risk of disease incidence and reduced life expectancy[38], these communities are under-represented in medical research[39], a situation that requires urgent remedy.

Physical measures were not included in The Lancet commission's life course prevention model[26], but their component factors have all been linked to dementia in previous observational studies, highlighting their importance. Psychosocial factors and local environment accounted for a relatively smaller portion of dementia cases. The low PAF associated with psychosocial factors may be because only emotional factors and isolation were included, with depression sited in the medical history category. No studies have explored the association between overall mood status and dementia, but personality is known to affect dementia risk[40].

Our estimation of the PAF of potentially modifiable risk factors is higher than previous estimates. The additional prevention contribution might come from those relatively unexplored factors as we, for the first time, undertook a hypothesis-free design and simultaneously evaluated a wide range of potential risk factors for dementia. Considering the co-occurrence of risk factors in one individual, the PAFs could also be inflated due to underlying correlation. Although we excluded factors with high collinearity and used weighting of the PAF for each domain to account for their non-independence, which is a more conservative strategy[41], the inter-relations might be more complex than a simple assessment of co-occurring risk factors[42]. Despite the limitation, the results are optimistic, indicating that dementia might be more preventable than previously thought.

Among the modifiable risk factors identified, it is necessary to pinpoint that some factors are modifiable for dementia prevention at an individual level including lifestyles, medical history, psychosocial factors and physical measures. Lifestyles incorporating cognitive activity, physical activity, smoking and alcohol drinking were proven to be associated with memory decline[43] and life expectancy[44] in 'old' population. Surgical or pharmacal interventions for individuals suffering from diseases like cataract and depression could reverse the disease-predicted risk of dementia[45,46]. Further, maintaining a good condition of psychical and physical health via healthier lifestyles has been proposed as suggestions for evidence-based dementia prevention guideline[47]. However, the other factors involving local environment and SES are potentially modifiable at a community level. Considering the group effect of these factors, it is urgent to call on government or other public organizations to make efforts in managing air pollution and natural environment, and promoting employment and education especially for low- and middle-income countries.

In our stratum-specific analyses, about 25% of the risk factors show potential direction shifts in different age groups or follow-up durations, regardless of the association significance. However, nearly all the 62 factors included in our PAF calculations did not show direction shifts in stratum analyses (except for 'Daytime dozing/sleeping (narcolepsy) _often'). Therefore, we believe that our primary conclusions were not affected by the potential changes in association direction in the long preclinical dementia period.

A substantial number of associations were not replicated in the MR analysis. Possible reasons include reverse causation and insufficient instrumental strength[48]. Inconsistent results do not necessarily negate the significance of risk factors; rather, additional confirmation is required. Estimates of MR analysis reflect lifelong average effects of genetic variants on disease, which may have contributed to the discrepancies if longitudinal analysis had been conducted over a shorter time period.

The strengths of this study include the large sample size, extensive exposure entries and data completeness provided by UK Biobank, which allowed us to systematically explore the modifiable risk factors. Moreover, we performed presumably the largest EWAS and examined the synergistic effect of related exposures, which would add knowledge to dementia prevention in an era of big data.

Our study has limitations. First, a high level of volunteer participation bias might confine our interpretation of results. In addition, it has a relatively young age at recruitment and a short follow-up duration. Consequently, it has a low incidence of dementia. However, valid assessment of exposure-disease relationships may be widely generalizable and participants representative of the population at large was not necessary for this purpose[49]. A recent study also found similar directions of risk associations between UKB and 18 representative studies, increasing the credibility of the generalizability of these results[50]. Second, our PAF estimates may not apply to other populations since PAF is associated with epidemiological data for each country, region and ethnicity[51,52]. Although risk factors in the UK Biobank was proven

to have comparable hazard ratios to those of representative studies[50], and we prevented the effect of unrepresentative prevalence data for UKB risk factors on PAF estimation by mitigating the most severe 1/3 or 2/3 of risk factors in the population, we acknowledge that these estimations may change as more cohort data from other regions become available. Validation in other populations is needed before implementation as public health strategies. Third, we were limited by the quality of variables measured by UKB, including both risk factors and dementia occurrence. Most risk factors were self-reported, and the diagnosis of dementia in UKB was mainly registry-based, lacking a comprehensive assessment with neuropsychological testing. Fourth, Bonferroni corrections are often over-conservative and might conceal possible associations that would be significant if studied alone. However, we manually removed meaningless ambiguous entries that might mask other meaningful variables, mitigating against inflated levels of significance due to multiple comparisons. Furthermore, the conclusion of causal relationships drawn from observational studies could be weakened or interpreted by other factors. For instance, insomnia decreases the risk of dementia in longitudinal analysis, but this association might be a proxy for a highly demanding job[3]. To reduce the effect of the interaction of factors on the results, we examined the collinearity of the factors and mutually adjusted different domains in subsequent analyses, but potential residual confounding might still exist. To understand the true associations, hypothesis-driven clinical trials or longitudinal studies with long follow-ups and a wide range of covariates are recommended.

## Conclusion

Dementia is associated with multidomain risk factors. Worsening profiles of physical measures, lifestyles, medical history, SES, psychosocial factors and local environment were related to an increased risk of dementia. Novel factors identified in EWAS are promising candidates for future clinical trials, although further confirmation of protective exposure levels is needed before implementation.

The estimation of PAF from our longitudinal study highlights the priority of interventions or changes to lifestyles that promote cardiovascular health and mental well-being as well as effective management of diabetes. Further reductions in risk can be achieved by oral and dental hygiene and treatment, resolution of gastrointestinal problems and support for physical disabilities. Unsurprisingly, low SES is a major risk factor for dementia and should be given prominence in public health policymaking in the context of an ageing global population.

## Methods

### Study design

We first explored the associations between 210 modifiable risk factors and incident dementia in an EWAS applying Cox proportional hazard regression models (Fig. 1). Factors associated with incident dementia (threshold for Bonferroni-corrected $P < 1.87 \times 10^{-4}$) were further validated in two-sample MR analyses. The joint effects of variables were evaluated by establishing weighted standardized scores for each domain, and their independent influence on dementia was assessed in a multivariable model and further stratified by genetic risk. The corresponding PAF was generated considering the non-independence (that is, co-occurrence of risks in the same individual) of different domains.

### Study population and dementia diagnosis

The present analyses used data from the UKB study, which recorded the baseline data of over 500,000 participants from 2006 to 2010 in the United Kingdom[53]. Participants were followed up until the date of first diagnosis, death, loss to follow-up or the last date with available information (December 2020), whichever came first. The dementia diagnoses were ascertained according to the corresponding three-character ICD codes (F00, F01, F02, F03, G30), extracted from UKB health outcome datasets first occurrences of health outcomes (Category 1712, including cases from hospital record, death registration and primary care) and algorithmically defined outcomes (Category 42). All participants provided written consent and approval was given by the North West Multi-centre Research Ethics Committee (MREC, https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics). Participants withdrawn from UKB as of December 2020, those with dementia at baseline, those with more than 20% missing values and those without available data on outcome and covariates were excluded (Fig. 1). The present analyses were conducted under UKB application number 19542.

### Modifiable factors

We excluded variables with more than 20% missing values and those that could not be modified (exploring variables at https://biobank.ndph.ox.ac.uk/showcase/browse.cgi). Then, we collected data on the remaining 210 modifiable factors that were measured or derived at baseline (Supplementary Tables 1–4). These variables were divided into 7 categories: (1) medical history (for example, disability), (2) early life factors (for example, breastfed as a baby), (3) local environment (for example, air pollution), (4) lifestyles (for example, alcohol intake), (5) psychosocial factors (for example, irritability), (6) physical measures (for example, forced vital capacity), and (7) socioeconomic status (SES, for example, household income).

To analyse the relative effects of all available exposures, potentially related variables (for example, 16 h and 24 h noise pollution) were preserved[15]. The Supplementary Note and Supplementary Tables 1–4 detail data cleansing and processing. Data were imputed using the multiple imputation by chained equations approach, with 5 imputed datasets and 10 iterations. We specified a predictive mean matching model for each variable, including the 11 most correlated predictors (sex was required to be included; Supplementary Note)[54].

### Statistical analyses

To conduct the EWAS, Cox proportional hazard regression models were applied to test the associations between each baseline exposure and incident dementia, with a conservative Bonferroni-corrected significance threshold for identifying top hits (0.05 divided by 268 tests, or $1.87 \times 10^{-4}$). These associations were adjusted for baseline age, sex, *APOE* ε4 status and assessment centre. An interaction with time of follow-up was added if the hazard proportional assumption was violated (a test using Schoenfeld's residuals had a $P < 0.0001$)[54]. We then checked the collinearity and excluded one of two factors from a highly associated pair ($r^2 > 0.9$; Supplementary Note). In sensitivity analyses, we used a machine learning method (hierarchical clustering) to alleviate collinearity (Supplementary Note). Stratum-specific analyses were conducted according to age at baseline (<65 and ≥65 yr), sex (male and female)[55], *APOE* ε4 status[4] and follow-up time, also utilizing the Bonferroni-correction method to determine top hits. Sex and *APOE* ε4 status were excluded as covariates in the subgroups. Two-sample MR analyses were adopted to further probe the findings in the EWAS. For risk factors, we searched the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/) to acquire publicly available UKB-based summary statistics. A GWAS for dichotomized sleep duration was performed to be consistent with our categorization strategy. IGAP meta-analysed GWAS datasets consisted of 21,982 Alzheimer's disease (AD) cases and 41,944 cognitively normal controls[56], which to our knowledge do not include a large number of samples from the UKB dataset. Summary statistics for all-cause dementia from FinnGen releases (https://r8.finngen.fi/pheno/F5_DEMENTIA) were also used for consistency with previous correlation analysis. We extracted significant single nucleotide polymorphisms for instrument variables ($P < 5 \times 10^{-8}$). For traits lacking single nucleotide polymorphisms (≤3) after outliers removal[15], we relaxed the threshold ($P < 5 \times 10^{-6}$) or excluded the traits[5]. We conducted MR analysis using the TwoSampleMR package in R[57]. We primarily performed the inverse-variance weighted method.

Potential heterogeneity and horizontal pleiotropy were examined by Cochran's *Q* test, MR Egger intercept and MR-PRESSO (pleiotropy residual sum and outlier) global test. Results were generated after outliers removal[15]. Other MR methods including weighted median[58] and Mendelian randomization-Egger (MR Egger) regression[59] were performed to generate robust conclusions.

Variables associated with dementia in the EWAS were further divided into six domains: medical history, lifestyles, SES, physical measures, psychosocial factors and local environment. Protective factors in the EWAS (HR < 1) were reverse coded to indicate the detrimental aspect, and participants scored 1 point for each detrimental factor (HR > 1). Weighted standardized scores for each domain were generated on the basis of $\beta$ coefficients of each variable in the Cox models[60], with risk factors mutually adjusted (belonging to one domain) and with adjustment for age, sex, *APOE* ε4 status and assessment centre. The original binary variables were multiplied by the $\beta$ coefficients, summed and divided by the sum of the $\beta$ coefficients. A higher score indicated exposure to more risk factors. We further divided the scores into tertiles as favourable (lower risk), intermediate or unfavourable (higher risk).

A Cox model was used to examine the associations of categorized lifestyles, medical history, local environment, psychosocial factors, physical measures and SES with incident dementia, adjusted for age, sex, *APOE* ε4 status and assessment centre (Model 1). A second model added all six domains mutually adjusted (Model 2). Genetic risks were evaluated using the *APOE* genotype, with the presence of ε4 indicating a high genetic risk, ε3ε3 an intermediate risk and others a low risk[4]. Interactions between the six domains and genetic risk were first tested by adding the cross-product terms of the continuous composite scores for the six domains and the genetic risk score, then stratified analyses were performed within genetic risk categories. We set follow-up time to more than 6 yr to reduce the risk of reverse causality. The proportionality of hazards assumption was assessed using the Schoenfeld residuals technique. We performed the same analysis using unweighted scores in sensitivity analyses.

The PAF represents the proportion of disease reduction that would occur if a given risk factor was exchanged for a more favourable alternative. It is an essential measure for forming public health and policy programmes. Estimating PAF is more straightforward for categorical variables (especially binary variables that avoid ambiguity in interpretation)[20]. Therefore, we combined the intermediate and favourable profiles of six domains to generate more conservative results (Model 1, eliminating the worst 1/3 of risk factors). A more complete elimination of risk factors by combining intermediate and unfavourable profiles (Model 2, eliminating the worst 2/3 of risk factors) was also tested. We first generated PAF for each domain on the basis of data provided by UK Biobank using the stdReg R package[61] in univariate logistic regression models adjusted for age, sex and *APOE* ε4 status. Follow-up time was fixed to at least 6 yr. Considering the non-independence of the six domains, communality was then calculated through principal component analysis to estimate the weight of each PAF, which was then used to compute both combined weighted PAF and individual weighted PAF[26]. Such method takes into account the co-occurrence of risk factors in the same individual and thus reduces the overestimation of PAF caused by the interactions of factors[62].

All *P* values were two-sided and analyses were conducted using R v 4.0.3 and Python 3.9.

### Reporting summary

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

## Data availability

The data used in the present study are available from UKB with restrictions applied. Data were used under license and are thus not publicly available. Access to the UKB data can be requested through a standard protocol (https://www.ukbiobank.ac.uk/register-apply/). Publicly available UKB-based summary statistics for the GWAS of risk factors can be obtained from the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/). The summary statistics of AD GWAS can be accessed from https://gwas.mrcieu.ac.uk/datasets/ieu-b-2/. The summary statistics of all-cause dementia GWAS can be accessed from https://r8.finngen.fi/pheno/F5_DEMENTIA.

## Code availability

Scripts used to perform the analyses are available at https://github.com/atticatto/UKB_AD_EWAS.git.

## References


1. Grande, G., Qiu, C. & Fratiglioni, L. Prevention of dementia in an ageing world: evidence and biological rationale. *Ageing Res. Rev.* **64**, 101045 (2020).
2. Kivipelto, M., Mangialasche, F. & Ngandu, T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nat. Rev. Neurol.* **14**, 653–666 (2018).
3. Palpatzis, E., Bass, N., Jones, R. & Mukadam, N. Longitudinal association of apolipoprotein E and sleep with incident dementia. *Alzheimers Dement*. https://doi.org/10.1002/alz.12439 (2021).
4. Licher, S. et al. Genetic predisposition, modifiable-risk-factor profile and long-term dementia risk in the general population. *Nat. Med.* **25**, 1364–1369 (2019).
5. Andrews, S. J., Fulton-Howard, B., O'Reilly, P., Marcora, E. & Goate, A. M. Causal associations between modifiable risk factors and the Alzheimer's phenome. *Ann. Neurol.* **89**, 54–65 (2021).
6. Pathan, S. S. et al. Association of lung function with cognitive decline and dementia: the Atherosclerosis Risk in Communities (ARIC) Study. *Eur. J. Neurol.* **18**, 888–898 (2011).
7. Chen, R. Association of environmental tobacco smoke with dementia and Alzheimer's disease among never smokers. *Alzheimers Dement*. **8**, 590–595 (2012).
8. Gardner, R. C. et al. Dementia risk after traumatic brain injury vs nonbrain trauma: the role of age and severity. *JAMA Neurol.* **71**, 1490–1497 (2014).
9. Lee, K. H. & Choi, Y. Y. Association between oral health and dementia in the elderly: a population-based study in Korea. *Sci. Rep.* **9**, 14407 (2019).
10. Duchowny, K. A. et al. Associations between handgrip strength and dementia risk, cognition, and neuroimaging outcomes in the UK Biobank cohort study. *JAMA Netw. Open* **5**, e2218314 (2022).
11. Korologou-Linden, R. et al. The causes and consequences of Alzheimer's disease: phenome-wide evidence from Mendelian randomization. *Nat. Commun.* **13**, 4726 (2022).
12. Johannesdottir Schmidt, S. A., Veres, K., Sørensen, H. T., Obel, N. & Henderson, V. W. Incident herpes zoster and risk of dementia: a population-based Danish cohort study. *Neurology* https://doi.org/10.1212/wnl.0000000000200709 (2022).
13. Lin, B. D. et al. Nongenetic factors associated with psychotic experiences among UK Biobank participants: exposome-wide analysis and Mendelian randomization analysis. *JAMA Psychiatry* **79**, 857–868 (2022).
14. Patel, C. J. & Ioannidis, J. P. Studying the elusive environment in large scale. *JAMA* **311**, 2173–2174 (2014).
15. Choi, K. W. et al. An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. *Am. J. Psychiatry* **177**, 944–954 (2020).
16. Patel, C. J., Bhattacharya, J., Ioannidis, J. P. A. & Bendavid, E. Systematic identification of correlates of HIV infection: an X-wide association study. *AIDS* **32**, 933–943 (2018).
17. Sheehan, A., Freni Sterrantino, A., Fecht, D., Elliott, P. & Hodgson, S. Childhood type 1 diabetes: an environment-wide association study across England. *Diabetologia* **63**, 964–976 (2020).

18. Manrai, A. K. et al. Informatics and data analytics to support exposome-based discovery for public health. *Annu Rev. Public Health* **38**, 279–294 (2017).
19. Zhuang, X. et al. Toward a panoramic perspective of the association between environmental factors and cardiovascular disease: an environment-wide association study from National Health and Nutrition Examination Survey 1999–2014. *Environ. Int.* **118**, 146–153 (2018).
20. Ritchie, K. et al. Designing prevention programmes to reduce incidence of dementia: prospective cohort study of modifiable risk factors. *Brit. Med. J.* **341**, c3885 (2010).
21. Grande, G., Ljungman, P. L. S., Eneroth, K., Bellander, T. & Rizzuto, D. Association between cardiovascular disease and long-term exposure to air pollution with the risk of dementia. *JAMA Neurol.* **77**, 801–809 (2020).
22. Bunch, T. J. Atrial fibrillation and dementia. *Circulation* **142**, 618–620 (2020).
23. Biessels, G. J. & Despa, F. Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications. *Nat. Rev. Endocrinol.* **14**, 591–604 (2018).
24. Singh-Manoux, A. et al. Trajectories of depressive symptoms before diagnosis of dementia: a 28-year follow-up study. *JAMA Psychiatry* **74**, 712–718 (2017).
25. Hu, X., Wang, T. & Jin, F. Alzheimer's disease and gut microbiota. *Sci. China Life Sci.* **59**, 1006–1023 (2016).
26. Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446 (2020).
27. Huang, S. Y. et al. Sleep, physical activity, sedentary behavior, and risk of incident dementia: a prospective cohort study of 431,924 UK Biobank participants. *Mol. Psychiatry* **27**, 4343–4354 (2022).
28. Wang, J. et al. Poor pulmonary function is associated with mild cognitive impairment, its progression to dementia, and brain pathologies: a community-based cohort study. *Alzheimers Dement.* https://doi.org/10.1002/alz.12625 (2022).
29. Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U. & Tucker-Drob, E. M. Education and cognitive functioning across the life span. *Psychol. Sci. Public Interest* **21**, 6–41 (2020).
30. Pan, X., Luo, Y. & Roberts, A. R. Secondhand smoke and women's cognitive function in China. *Am. J. Epidemiol.* **187**, 911–918 (2018).
31. Ma, L. Z. et al. Time spent in outdoor light is associated with the risk of dementia: a prospective cohort study of 362094 participants. *BMC Med.* **20**, 132 (2022).
32. Calvin, C. M., Conroy, M. C., Moore, S. F., Kuzma, E. & Littlejohns, T. J. Association of multimorbidity, disease clusters, and modification by genetic factors with risk of dementia. *JAMA Netw. Open* **5**, e2232124 (2022).
33. Tai, X. Y. et al. Cardiometabolic multimorbidity, genetic risk, and dementia: a prospective cohort study. *Lancet Healthy Longev.* **3**, e428–e436 (2022).
34. Baumgart, M. et al. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: a population-based perspective. *Alzheimers Dement.* **11**, 718–726 (2015).
35. Yaffe, K. et al. Effect of socioeconomic disparities on incidence of dementia among biracial older adults: prospective study. *Brit. Med. J.* **347**, f7051 (2013).
36. Holm, E. et al. Frequency of missed or delayed diagnosis in dementia is associated with neighborhood socioeconomic status. *Alzheimers Dement.* **8**, e12271 (2022).
37. Chan, M. Y. et al. Socioeconomic status moderates age-related differences in the brain's functional network organization and anatomy across the adult lifespan. *Proc. Natl Acad. Sci. USA* **115**, e5144–e5153 (2018).
38. Marmot, M. Social determinants of health inequalities. *Lancet* **365**, 1099–1104 (2005).
39. Heiat, A., Gross, C. P. & Krumholz, H. M. Representation of the elderly, women, and minorities in heart failure clinical trials. *Arch. Intern. Med.* **162**, 1682–1688 (2002).
40. Low, L. F., Harrison, F. & Lackersteen, S. M. Does personality affect risk for dementia? A systematic review and meta-analysis. *Am. J. Geriatr. Psychiatry* **21**, 713–728 (2013).
41. Livingston, G. et al. Dementia prevention, intervention, and care. *Lancet* **390**, 2673–2734 (2017).
42. Norton, S., Matthews, F. E., Barnes, D. E., Yaffe, K. & Brayne, C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol.* **13**, 788–794 (2014).
43. Jia, J. et al. Association between healthy lifestyle and memory decline in older adults: 10 year, population based, prospective cohort study. *Brit. Med. J.* **380**, e072691 (2023).
44. Dhana, K. et al. Healthy lifestyle and life expectancy with and without Alzheimer's dementia: population based cohort study. *Brit. Med. J.* **377**, e068390 (2022).
45. Yang, L. et al. Depression, depression treatments, and risk of incident dementia: a prospective cohort study of 354,313 Participants. *Biol. Psychiatry* https://doi.org/10.1016/j.biopsych.2022.08.026 (2022).
46. Ma, L. Z. et al. Cataract, cataract surgery, and risk of incident dementia: a prospective cohort study of 300,823 participants. *Biol. Psychiatry* https://doi.org/10.1016/j.biopsych.2022.06.005 (2022).
47. Yu, J. T. et al. Evidence-based prevention of Alzheimer's disease: systematic review and meta-analysis of 243 observational prospective studies and 153 randomised controlled trials. *J. Neurol. Neurosurg. Psychiatry* **91**, 1201–1209 (2020).
48. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
49. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
50. Batty, G. D., Gale, C. R., Kivimäki, M., Deary, I. J. & Bell, S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *Brit. Med. J.* **368**, m131 (2020).
51. Lee, M. et al. Variation in population attributable fraction of dementia associated with potentially modifiable risk factors by race and ethnicity in the US. *JAMA Netw. Open* **5**, e2219672 (2022).
52. Ma'u, E., Cullum, S., Cheung, G., Livingston, G. & Mukadam, N. Differences in the potential for dementia prevention between major ethnic groups within one country: a cross sectional analysis of population attributable fraction of potentially modifiable risk factors in New Zealand. *Lancet Reg. Health West Pac.* **13**, 100191 (2021).
53. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
54. Ganna, A. & Ingelsson, E. 5 year mortality predictors in 498103 UK Biobank participants: a prospective population-based study. *Lancet* **386**, 533–540 (2015).
55. Ferretti, M. T. et al. Sex differences in Alzheimer disease - the gateway to precision medicine. *Nat. Rev. Neurol.* **14**, 457–469 (2018).
56. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
57. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* https://doi.org/10.7554/eLife.34408 (2018).

58. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
59. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
60. Lourida, I. et al. Association of lifestyle and genetic risk with incidence of dementia. *JAMA* **322**, 430–437 (2019).
61. Sjölander, A. Estimation of causal effect measures with the R-package stdReg. *Eur. J. Epidemiol.* **33**, 847–858 (2018).
62. Kivipelto, M. & Mangialasche, F. Alzheimer disease: to what extent can Alzheimer disease be prevented? *Nat. Rev. Neurol.* **10**, 552–553 (2014).

## Acknowledgements

J.-T.Y. was supported by grants from the Science and Technology Innovation 2030 Major Projects (2022ZD0211600); the National Natural Science Foundation of China (82071201, 81971032, 92249305); 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); Shanghai Talent Development Funding for The Project (2019074); ZHANGJIANG LAB, Tianqiao and Chrissy Chen Institute; and the State Key Laboratory of Neurobiology and Frontiers Centre for Brain Science of the Ministry of Education, Fudan University. W.C. was supported by grants from the National Natural Sciences Foundation of China (no. 82071997) and the Shanghai Rising-Star Program (no. 21QA1408700). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank all the participants and professionals contributing to the UK Biobank; the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses; and the participants and investigators of the FinnGen study. 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. 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/PERADES was supported by the Medical Research Council (Grant no. 503480), Alzheimer's Research UK (Grant no. 503176), the Wellcome Trust (Grant no. 082604/2/07/Z) and the German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant no. 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by NIH/NIA grant R01 AG033193 and NIA AG081220, and AGES contract N01–AG–12100, NHLBI grant R01 HL105756, the Icelandic Heart Association, the Erasmus Medical Center and Erasmus University. ADGC was supported by NIH/NIA grants U01 AG032984, U24 AG021886 and U01 AG016976, and Alzheimer's Association grant ADGC–10–196728.

## Author contributions

J.-T.Y. and W.C. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. J.-T.Y. conceived and designed the project. All authors acquired, analysed or interpreted data. Y.Z., S.-D.C., Y.-T.D., A.D.S., J.S. and J.-T.Y. wrote the initial draft of the manuscript. Y.Z., S.-D.C., Y.-T.D., J.Y., X.-Y.H., X.-R.W., B.-S.W., L.Y., Y.-R.Z., K.K., A.D.S., J.S., W.C. and J.-T.Y. critically revised the manuscript for important intellectual content. Y.Z., S.-D.C., Y.-T.D. and J.Y. conducted statistical analysis. J.-F.F., W.C. and J.-T.Y. acquired funding. J.-F.F., W.C. and J.-T.Y. provided administrative, technical or material support. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s41562-023-01585-x.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01585-x.

**Correspondence and requests for materials** should be addressed to Jin-Tai Yu.

**Peer review information** *Nature Human Behaviour* thanks Marios Georgakis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

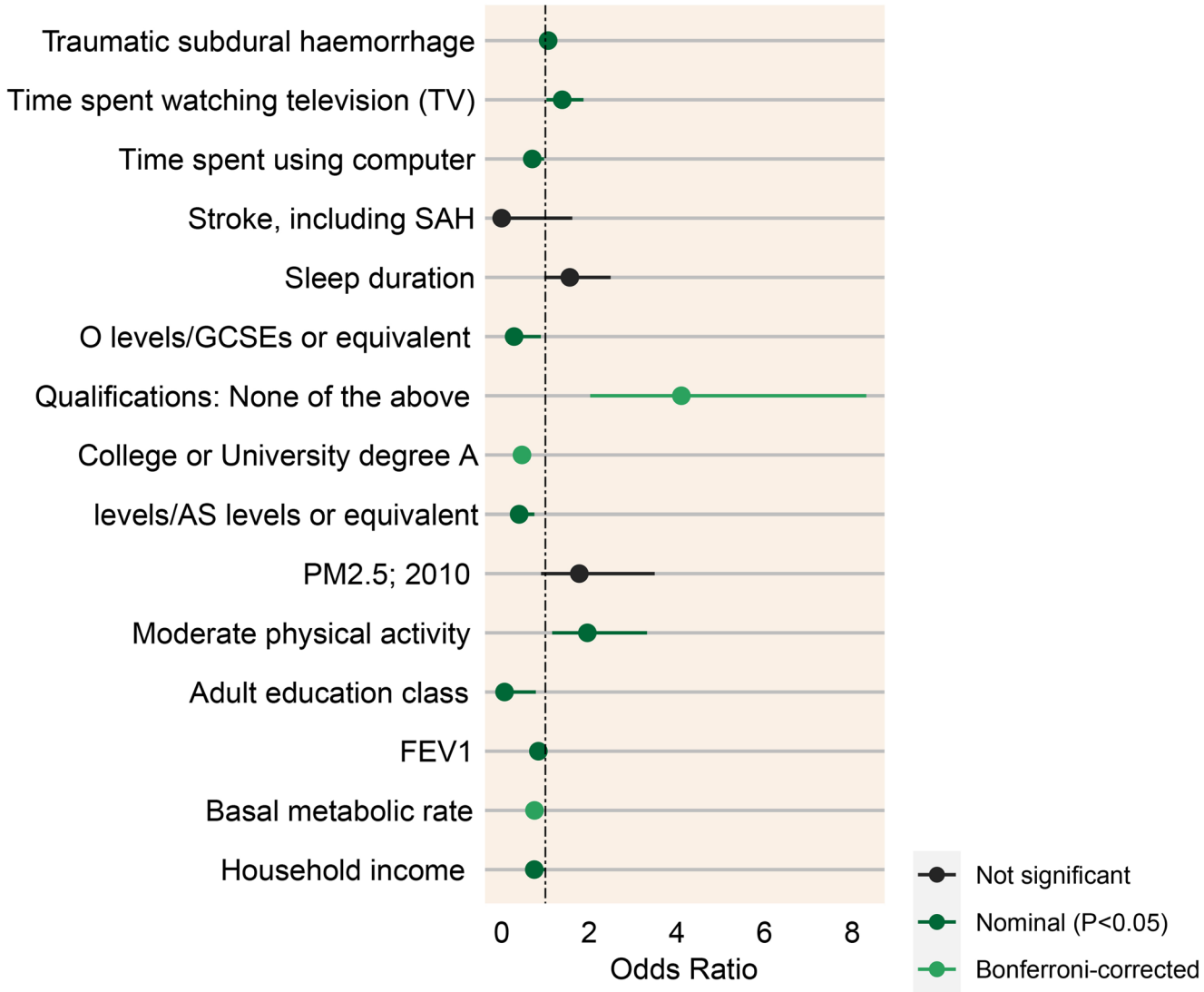


UK Biobank has received ethical approval from the North West Multi-centre Research Ethics Committee (MREC, https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics), and informed consent through electronic signature was obtained from study participants. This study utilized the UK Biobank Resource under application number 19542.

https://doi.org/10.1038/s41562-023-01585-x


**Extended Data Fig. 1 | Associations between modifiable risk factors and incident dementia when excluding participants who developed dementia within the first 5 years or 10 years.** Models were adjusted for baseline age, sex, *APOE* ε4 status, and assessment center. The color of cells indicates the effect sizes (HR) between each risk factor and incident dementia (N = 344,324). Asterisks in cells represent significant associations after correction for multiple testing (Bonferroni-corrected, $P < 1.87 \times 10^{-4}$).

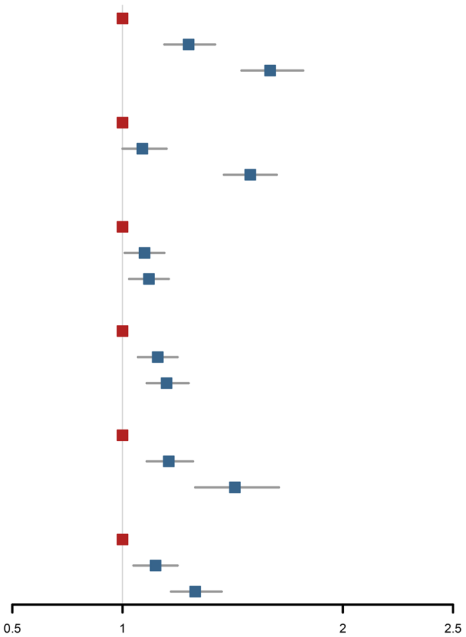
https://doi.org/10.1038/s41562-023-01585-x




**Extended Data Fig. 2 | Mendelian randomization estimates of factors in relation to dementia risk.** Estimates were generated using inverse-variance weighted method after removing outliers. Results generated using other methods are available in Supplementary Table 11. Dots represent odds ratios and lines represent 95% CIs.

https://doi.org/10.1038/s41562-023-01585-x

| | Total No. of participants | No. of Cases of Dementia | HR (95% CI) | P Value | P for trend |
|---|---|---|---|---|---|
| **Lifestyle** | | | | | |
| Favorable | | | 1 (Reference) | | |
| Intermediate | 97597 | 1211 | 1.30 (1.19–1.42) | $4.19\times10^{-9}$ | |
| Unfavorable | 100467 | 1842 | 1.67 (1.54–1.82) | $< 2\times10^{-16}$ | $< 2\times10^{-16}$ |
| **Medical history** | | | | | |
| Favorable | | | 1 (Reference) | | |
| Intermediate | 51078 | 756 | 1.09 (1.00–1.20) | 0.058 | |
| Unfavorable | 97557 | 1984 | 1.58 (1.46–1.70) | $< 2\times10^{-16}$ | $< 2\times10^{-16}$ |
| **Local environment** | | | | | |
| Favorable | | | 1 (Reference) | | |
| Intermediate | 86589 | 1222 | 1.10 (1.01–1.19) | 0.022 | |
| Unfavorable | 99873 | 1365 | 1.12 (1.03–1.21) | $7.10\times10^{-3}$ | 0.006 |
| **Psychosocial factors** | | | | | |
| Favorable | | | 1 (Reference) | | |
| Intermediate | 90773 | 1231 | 1.16 (1.07–1.25) | 0.0003 | |
| Unfavorable | 95426 | 1330 | 1.20 (1.11–1.30) | $3.36\times10^{-6}$ | $1.36\times10^{-6}$ |
| **Physical measures** | | | | | |
| Favorable | | | 1 (Reference) | | |
| Intermediate | 98290 | 1339 | 1.21 (1.11–1.32) | $1.08\times10^{-5}$ | |
| Unfavorable | 96672 | 1267 | 1.51 (1.33–1.71) | $7.28\times10^{-11}$ | $9.87\times10^{-11}$ |
| **Socioeconomic status** | | | | | |
| Favorable | | | 1 (Reference) | | |
| Intermediate | 96575 | 1321 | 1.15 (1.05–1.25) | 0.003 | |
| Unfavorable | 90051 | 1773 | 1.33 (1.22–1.45) | $3.73\times10^{-10}$ | $2.85\times10^{-10}$ |




**Extended Data Fig. 3 | Associations between six domains and dementia based on factors selected by machine learning.** The favourable profile was set as reference in each domain. The associations were estimated applying Cox model including all six domains mutually adjusted and with adjustment of age, sex, *APOE* ε4 status, and assessment center. Dots represent hazard ratios; horizontal lines indicate corresponding 95% CIs. Z-tests were used to assess statistical significance and derive Z statistics and corresponding two-sided P values. HR, hazard ratio; CI, confidence interval. SES, socioeconomic status.

nature portfolio

Corresponding author(s): Jin-Tai Yu

Last updated by author(s): Mar 10, 2023

# Reporting Summary

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

## Statistics

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

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

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was involved in data collection (data used is all directly available from UK Biobank, as described in detail in the paper) |
| Data analysis | R version 4.0.3 packages:<br>survival 3.2 was used to perform Cox proportional hazard regression model in the exposure wide association analysis;<br>TwoSampleMR 0.5.6 was used to perform Mendelian randomization study;<br>MRPRESSO 1.0 for MR-PRESSO analysis;<br>mice 3.15.0 package was used for data imputation;<br>stdReg 3.4.1 was used to generate PAF;<br>psych 2.2.9 was used to generate tetrachoric correlation matrix in the calculation of communality.<br>Plink 1.90 and PRSice (v2.3.3) was used to for the generation of AD-PRS.<br>Python 3.9 with scikit-learn v1.2.1 was used to perform the machine learning process. |

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

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

The data used in the present study are available from UKB with restrictions applied. Data were used under license and are thus not publicly available. Access to the UKB data can be requested through a standard protocol (https://www.ukbiobank.ac.uk/register-apply/). Publicly available UKB-based summary statistics for risk factors' GWAS can be obtained from MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/). The summary statistics of AD GWAS can be accessed from https://gwas.mrcieu.ac.uk/datasets/ieu-b-2/. The summary statistics of all-cause dementia GWAS can be accessed from https://r8.finngen.fi/pheno/F5_DEMENTIA.

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | We took sex into considerations in our study and our findings could apply to both male and female. Sex (Field ID 31) in the UK Biobank was determined based on self-reporting data via questionaire, and all included 344 324 participants gave written informed consent for sharing of individual-level data. |
| Population characteristics | This study included 344 324 adults (186 720 [54.2%] female; mean [SD] age, 56.01 [8.04] years). Over a mean (SD) of 8.81 (2.85) years of follow-up, 4654 participants (1.35%) developed dementia, and the mean (SD) age at diagnoses were 72.8 (5.9) years. The baseline demographic data of participants is shown in eTable 5. We calculated descriptive statistics as mean (SD) for continuous variables and number (percentage) for categorical variables. |
| Recruitment | The UKB enrolled the participants aged 40-69 years between 2006 and 2010 for baseline assessments in 22 centers across the UK. The assessment visits comprised interviews and questionnaires covering lifestyles and health conditions, physical measures, biological samples, imaging, and genotyping. The database is linked to national health datasets, including primary care, hospital inpatient, death, and cancer registration data. |
| Ethics oversight | UK Biobank has received ethical approval from the North West Multi-centre Research Ethics Committee (MREC, https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics), and informed consent through electronic signature was obtained from study participants. This study utilized the UK Biobank Resource under application number 19542. |

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

# Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

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

# Life sciences study design

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

| | |
|---|---|
| Sample size | No statistical methods were used to predetermine sample sizes. All currently available sample in the UK Biobank were included. After excluding 28073 without follow up data/ with baseline dementia, 104806 with missing data in >20% of variables, 25273 with missing data on outcome and/or covariates, 17 withdrawn from the study, 344,324 out of 502,493 UKB subjects were eligible and were included for analysis. |
| Data exclusions | Participants withdrawn from UKB as of December 2020, participants with dementia at baseline, participants with more than 20% missing values, and those without available data on outcome and covariates were excluded. |
| Replication | All available data were used to maximize statistical power of the analysis therefore we did not repeat the analysis. |
| Randomization | Covariates including baseline age, sex, APOE ε4 status, and assessment center were adjusted in the study. |
| Blinding | Blinding was not applicable to this study as this study is observational. |

# Reporting for specific materials, systems and methods

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

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |