



# Supplementary Material: Sunshine Duration and Dementia: Mediating Role of Vitamin D, Calcium, and Accelerated Brain Aging

## Table of contents

Supplementary Methods.....	2
Table S1 Sample sizes used in different analyses.....	3
Table S2 International Classification of Disease codes used to ascertain dementia and its subtypes4	
Table S3 UDI and notes of variables in the UK Biobank repository.....	5
Table S4 Characteristics of genetic variants associated with dementia in the UK biobank. ....	7
Table S5 Absolute correlation of covariates for unweighted and weighted populations. ....	8
Table S6 The results of additive interactions.....	9
Table S7 The results of sensitivity analyses. ....	10
Table S8 The results of sensitivity analyses for the mediation analysis. ....	11
Fig. S1 The flowchart of participants selection in the study.....	12
Fig. S2 The directed acyclic graph for covariate selection.....	13



## Supplementary Methods

### Generalized propensity score

The generalized propensity score (GPS) method is the extension of the propensity score (PS) approach to continuous exposures, which was initially developed by Keisuke Hirano and Guido W. Imbens [1, 2]. The generalized propensity score is a balancing score based on the conditional density of the continuous exposure given the confounders [3, 4].

$$GPS = r(t, x)$$

$$r(t, x) = f_{T|X}(t|x)$$

Note:  $t$  is continuous exposure.  $x$  is the confounders.  $f()$  is probability density function.

There are three assumptions for the GPS framework, which are the consistency, positivity, and ignorability assumptions. The consistency assumption is also known as the no-interference or stable-unit-treatment-value assumption. Briefly, we assume that the potential outcome for a given observation is not affected by the exposure of any other unit, and that each exposure defines a unique outcome for each observation [5]. The positivity assumption is that the probability of receiving any value of treatment conditional on confounders is greater than zero. In addition, the ignorability assumption indicates that all information about possible confounders has been collected.

Our approach to causal inference includes four steps: 1) estimating the GPSs; 2) constructing a pseudo-population using GPSs; 3) assessing the covariate balance of the constructed pseudo-population; 4) performing the analysis in the pseudo-population.

There three methods with GPS to create pseudo-population, including matching by GPS, weighting by GPS, and including GPS as a covariate in the main model. GPS with weighting is the most common method chosen by researchers. The weighting approach involves using the inverse of (generalized) propensity score to weight the observations. In practice, it is often recommended to "stabilise" the weights in order to reduce the variance of the estimates, especially when the exposure is continuous. The stabilised inverse probability weights (IPW) were calculated from the inverse of the GPS using the following formula:

$$\text{Stabilized Weight} = \frac{f(x)}{f(x/v)}$$

In the model,  $x$  represents the exposure and  $v$  represents the covariates. The  $f(x)$  is the probability density function of the exposure. The  $f(x/v)$  is the probability density function of the exposure given the covariates, which are the GPSs. We truncated the lowest and highest 1% of the IPW weights to stabilize the estimates [6].

**Table S1.** Sample sizes used in different analyses.

<b>Analyses</b>	<b>Variables</b>	<b>Sample size</b>
	Association between sunshine duration and incident dementia	
Time-dependent Cox proportional hazard model	Sunshine duration, age, sex, household income, physical activity, Townsend deprivation index, time spent outdoors, residential area, percentage of greenspace at 300 m buffers, ambient temperature, disease history (cardiovascular disease, hypertension, diabetes, stroke, self-reported long-term illness disability or frailty, self-reported fair or poor health status), and air pollution.	499,727
	Mediation analyses	
Vitamin D	Vitamin D, sunshine duration, age, sex, household income, physical activity, Townsend deprivation index, time spent outdoors, residential area, percentage of greenspace at 300 m buffers, ambient temperature, disease history (cardiovascular disease, hypertension, diabetes, stroke, self-reported long-term illness disability or frailty, self-reported fair or poor health status), and air pollution.	401,595
Calcium	Calcium, sunshine duration, age, sex, household income, physical activity, Townsend deprivation index, time spent outdoors, residential area, percentage of greenspace at 300 m buffers, ambient temperature, disease history (cardiovascular disease, hypertension, diabetes, stroke, self-reported long-term illness disability or frailty, self-reported fair or poor health status), and air pollution.	401,595
Brain age	Brain MRI data, sunshine duration, age, sex, household income, physical activity, Townsend deprivation index, time spent outdoors, residential area, percentage of greenspace at 300 m buffers, ambient temperature, disease history (cardiovascular disease, hypertension, diabetes, stroke, self-reported long-term illness disability or frailty, self-reported fair or poor health status), and air pollution.	33,661

**Table S2.** International Classification of Disease codes used to ascertain dementia and its subtypes.

	<b>All-Cause Dementia</b>	<b>Alzheimer's Dementia</b>	<b>Vascular Dementia</b>
<b>ICD-9</b>	290.2, 290.3, 290.4, 291.2, 294.1, 331.0, 331.1, 331.2, 331.5	331.0	290.4
<b>ICD-10</b>	F00, F00.0, F00.1, F00.2, F00.9, F01, F01.0, F01.1, F01.2, F01.3, F01.8, F01.9, F02, F02.0, F02.1, F02.2, F02.3, F02.4, F02.8, F03, F05.1, F10.6, G30, G30.0, G30.1, G30.8, G30.9, G31.0, A81.0, G31.1, G31.8, I67.3	F00, F00.0, F00.1, F00.2, F00.9, G30, G30.0, G30.1, G30.8, G30.9	F01, F01.0, F01.1, F01.2, F01.3, F01.8, F01.9, I67.3

**Table S3.** UDI and notes of variables in the UK Biobank repository.

Variables	UDI	Notes
Age	21011	Age at recruitment.
Sex	31	Classification: female, male.
Ethnicity	21000	Ethnicity is an amalgam of sequential branching questions asked during the initial Assessment Centre visit as part of the touchscreen questionnaire.
BMI	23104	BMI was measured by impedance measurement. Classification: underweight, <18.5; normal, 18.5-25; overweight, 25-30; obese, ≥30.
Annual household income	738	Income was measured with screening question: “What is the average total income before tax received by your household?”. Classification: <18,000£; 18,000-30,999£; 31,000-51,999£; >52,000£; unknown.
Townsend deprivation	189	Townsend deprivation index was calculated based on the preceding national census output area. A Townsend deprivation score was computed for all participants using information about employment, car and home ownership, and household overcrowding. Classification: low, moderate, or high.
Residential area	20118	The classification is derived by combining each participant’s home postcode with data generated from the 2001 census from the Office of National Statistics, using the Geoconvert tool from Census Dissemination Unit.
Education	6138	Education was measured with screening question: “Which of the following qualifications do you have? (You can select more than one)”. Classification: college or university degree; any school degree; vocational qualifications or other.
Physical activity	22032	Physical activity was classified according to International Physical Activity Questionnaire (IPAQ). Classification: High: 1) vigorous-intensity activity on at least 3 days achieving a minimum Total physical activity of at least 1500 MET-minutes/week; 2) 7 or more days of any combination of walking, moderate-intensity or vigorous-intensity activities achieving a minimum Total physical activity of at least 3000 MET-minutes/week. Moderate: 1) 3 or more days of vigorous-intensity activity of at least 20 minutes per day; 2) 5 or more days of moderate-intensity activity and/or walking of at least 30 minutes per day; 3) 5 or more days of any combination of walking, moderate-intensity or vigorous-intensity activities achieving a minimum Total physical activity of at least 600 MET-minutes/week.
Smoking status	20116	Low: Those individuals who not meet criteria for Categories high or moderate are considered ‘low’. The current/past smoking status of the participant.
Alcohol intake	1558	Heavy alcohol intake was measured with screening question: “About how often do you drink alcohol?”. Classification: yes: daily or almost daily drinking alcohol; no: occasional or never drinking alcohol.
Time spent outdoors in winter	1050	In a typical DAY in summer, how many hours do you spend outdoors?
Time spent outdoors in summer	1060	In a typical DAY in winter, how many hours do you spend outdoors?

Greenspace\_300

24503

Residential green space was estimated using land use data obtained from the 2005 Generalized Land Use

Biomarkers for brain age

25000- 25768

Obtained from UK Biobank assessment center visit.

---

**Abbreviations:** UDI, the Unique Data Identifier.

**Table S4.** Characteristics of genetic variants associated with dementia in the UK biobank.

<b>Rs number</b>	<b>Chromosome</b>	<b>Gene</b>	<b>Major /Minor allele</b>	<b>Odds ratios (OR)<sup>a</sup></b>	<b>95% CI for OR</b>
rs10467994	15	SPPL2A	T/C	0.94	0.92–0.96
rs10808026	7	EPHA1	C/A	0.9	0.88–0.93
rs11218343	11	SORL1	T/C	0.8	0.75–0.85
rs12881735	14	SLC24A4	T/C	0.92	0.89–0.94
rs17125924	14	FERMT2	A/G	1.14	1.09–1.18
rs2632516	17	MIR142/TSP0AP1-AS1d	G/C	0.94	0.91–0.96
rs2830500	21	ADAMTS1	C/A	0.93	0.91–0.96
rs429358	19	APOE	T/C	3.32	3.20–3.45
rs593742	15	ADAM10	A/G	0.93	0.91–0.95
rs6024870	20	CASS4	G/A	0.88	0.85–0.92
rs62039712	16	WVOX	G/A	1.16	1.10–1.23
rs6733839	2	BIN1	C/T	1.2	1.17–1.23
rs7185636	16	IQCK	T/C	0.92	0.89–0.95
rs7295246	12	ADAMTS20	T/G	1.06	1.04–1.08
rs73223431	8	PTK2B	C/T	1.1	1.07–1.13
rs7920721	10	ECHDC3	A/G	1.08	1.05–1.11
rs7933202	11	MS4A2	A/C	0.89	0.87–0.92
rs9473117	6	CD2AP	A/C	1.09	1.06–1.12
rs10933431	2	INPP5D	C/G	0.91	0.88–0.94
rs12539172	7	NYAP1g	C/T	0.92	0.90–0.95
rs3740688	11	SPI1h	T/G	0.92	0.89–0.94
rs3752246	19	ABCA7	C/G	1.15	1.11–1.18
rs3851179	11	PICALM	C/T	0.88	0.86–0.90
rs4735340	8	NDUFAF6	T/A	0.94	0.92–0.96
rs4844610	1	CR1	C/A	1.17	1.13–1.21
rs9331896	8	CLU	T/C	0.88	0.85–0.90

<sup>a</sup> Calculated with respect to the minor allele.

**Table S5.** Absolute correlation of covariates for unweighted and weighted populations.

<b>Variables</b>	<b>Age</b>	<b>Sex</b>	<b>Household income</b>	<b>Time spent outdoors</b>	<b>Physical activity</b>	<b>Residential areas</b>	<b>Deprivation index</b>	<b>Temperature</b>	<b>Air pollution</b>	<b>Mean</b>
Unweighted	0.016	0.024	0.074	0.021	0.015	0.094	0.057	0.257	0.271	0.093
Weighted	0.005	0.008	0.021	0.006	0.005	0.026	0.005	0.024	0.002	0.012

**Table S6.** The results of additive interactions.

Group	Cases	RERI (95% CI)
Gender		
Female	133880	Ref
Male	135916	-0.01 (-0.03, 0.01)
Age		
Young	221872	Ref
Old	47924	-0.02 (-0.08, 0.03)
Household income		
Greater than 31,000	53718	Ref
18,000 to 30,999	67145	-0.01 (-0.03, 0.01)
Less than 18,000	148933	-0.01 (-0.02, 0.01)
Alcohol intake		
Never	17502	Ref
Occasional	55528	-0.01 (-0.05, 0.02)
Moderate	136264	-0.01 (-0.03, 0.01)
Heavy	60502	-0.01 (-0.05, 0.02)
Smoking status		
Never smoked	148469	Ref
Previous smoked	95497	-0.02 (-0.06, 0.02)
Current smoked	25830	-0.02 (-0.09, 0.04)
Time spent outdoors		
High	129251	Ref
Low	140545	-0.02 (-0.05, 0.02)
Physical activity		
Low	48766	Ref
Moderate	110306	-0.02 (-0.06, 0.03)
High	110724	-0.02 (-0.06, 0.03)
Greenspace		
Low	134889	Ref
High	134907	-0.01 (-0.05, 0.02)

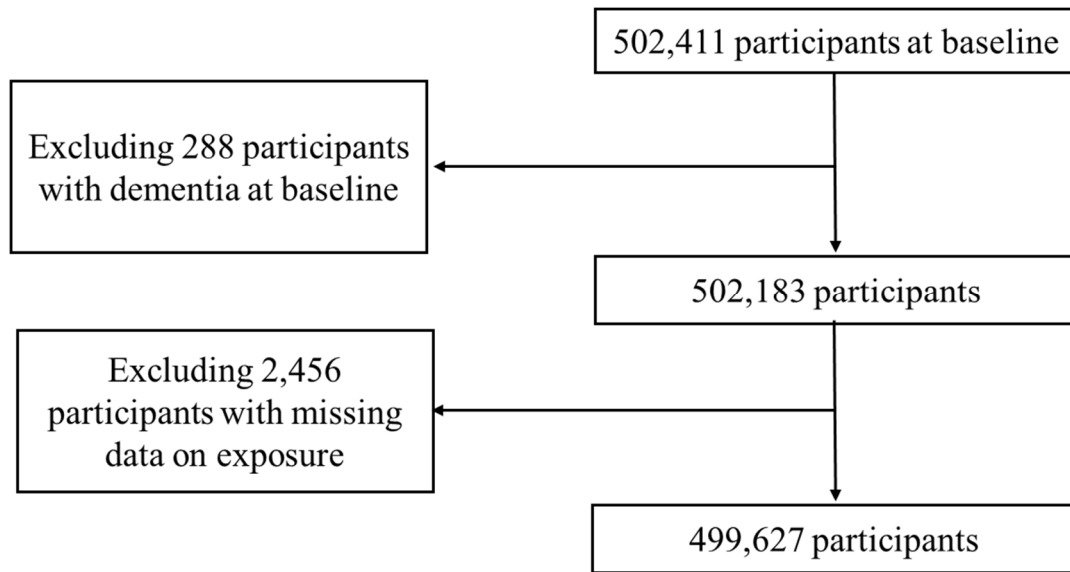
Table S7. The results of sensitivity analyses.

Model	All case dementia		Vascular dementia		Alzheimer dementia	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Multiple imputations with chained equations to impute missing data on covariates						
Model 1	1.08 (1.06, 1.11)	<0.001	1.27 (1.20, 1.34)	<0.001	1.08 (1.03, 1.12)	<0.001
Model 2	1.08 (1.06, 1.11)	<0.001	1.25 (1.15, 1.36)	<0.001	1.08 (1.04, 1.12)	<0.001
Model 3	1.05 (1.02, 1.08)	<0.001	1.18 (1.08, 1.28)	<0.001	1.05 (1.01, 1.09)	<0.001
Model 4	1.06 (1.03, 1.09)	<0.001	1.20 (1.09, 1.32)	<0.001	1.04 (1.01, 1.08)	0.030
Excluding participants with dementia events or death within 1 year of follow-up						
Model 1	1.08 (1.06, 1.11)	<0.001	1.27 (1.20, 1.34)	<0.001	1.08 (1.03, 1.12)	<0.001
Model 2	1.08 (1.06, 1.11)	<0.001	1.27 (1.20, 1.34)	<0.001	1.08 (1.04, 1.12)	<0.001
Model 3	1.05 (1.02, 1.08)	<0.001	1.20 (1.14, 1.27)	<0.001	1.05 (1.01, 1.09)	0.023
Model 4	1.06 (1.03, 1.09)	<0.001	1.26 (1.18, 1.34)	<0.001	1.04 (1.01, 1.08)	0.045
Including chronic diseases as covariates						
Model 1	1.07 (1.04, 1.10)	<0.001	1.26 (1.19, 1.33)	<0.001	1.07 (1.02, 1.11)	0.002
Model 2	1.05 (1.02, 1.08)	<0.001	1.22 (1.15, 1.29)	<0.001	1.04 (1.01, 1.09)	0.003
Model 3	1.04 (1.01, 1.07)	0.005	1.21 (1.14, 1.28)	<0.001	1.06 (1.02, 1.11)	0.003
Model 4	1.05 (1.02, 1.09)	0.001	1.26 (1.18, 1.35)	<0.001	1.06 (1.02, 1.11)	0.004
Including vitamin D intake as covariates						
Model 1	1.08 (1.06, 1.11)	<0.001	1.27 (1.20, 1.34)	<0.001	1.08 (1.04, 1.12)	<0.001
Model 2	1.06 (1.03, 1.09)	<0.001	1.23 (1.16, 1.30)	<0.001	1.05 (1.01, 1.10)	<0.001
Model 3	1.05 (1.02, 1.08)	<0.001	1.21 (1.15, 1.28)	<0.001	1.05 (1.01, 1.09)	<0.001
Model 4	1.07 (1.03, 1.10)	<0.001	1.27 (1.19, 1.35)	<0.001	1.05 (1.01, 1.10)	0.026
Using educational level as a surrogate for household income						
Model 1	1.08 (1.06, 1.11)	<0.001	1.27 (1.20, 1.34)	<0.001	1.08 (1.03, 1.12)	<0.001
Model 2	1.06 (1.03, 1.09)	<0.001	1.27 (1.20, 1.34)	<0.001	1.08 (1.04, 1.12)	<0.001
Model 3	1.05 (1.02, 1.08)	<0.001	1.23 (1.16, 1.30)	<0.001	1.05 (1.01, 1.10)	0.010
Model 4	1.07 (1.03, 1.10)	<0.001	1.27 (1.19, 1.35)	<0.001	1.05 (1.01, 1.09)	0.026
E-value						
Model 1	1.279		1.513		1.279	
Model 2	1.243		1.486		1.279	
Model 3	1.279		1.486		1.279	
Model 4	1.279		1.615		1.459	

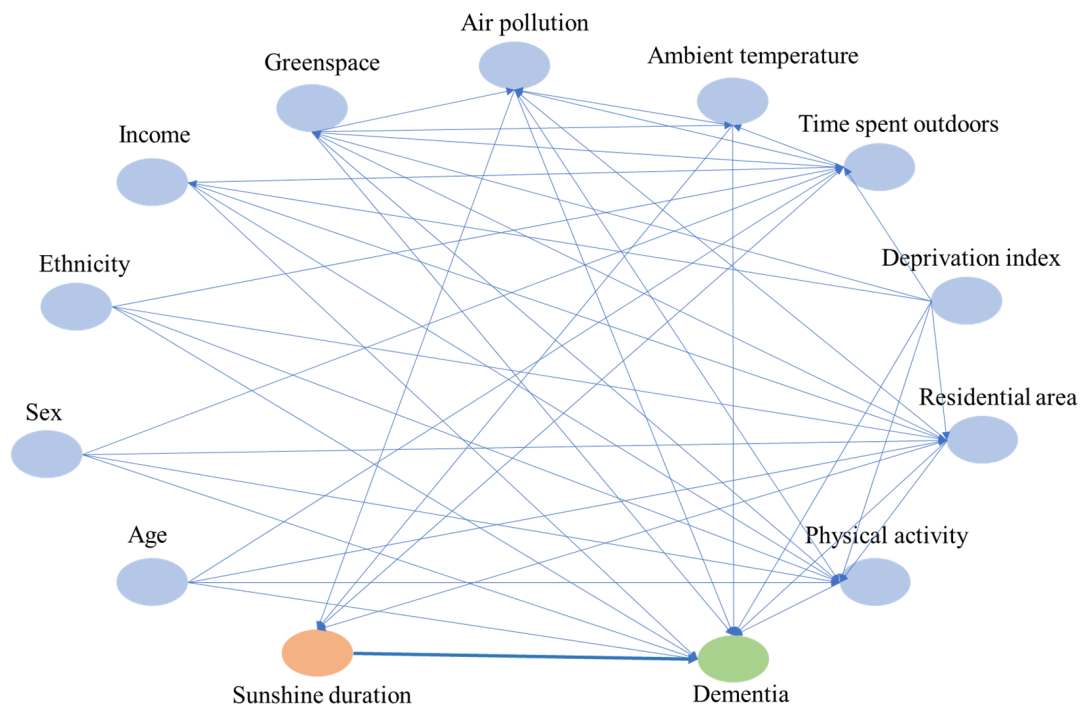
Note: Model 1 adjusted for age and sex. Model 2 additionally included adjustments for household income, physical activity, and Townsend deprivation index. Model 3 further incorporated adjustments for time spent outdoors, residential area, percentage of greenspace (%) at 300 m buffers, ambient temperature, and air pollution. Model 4 was the IPW-weighted model.

**Table S8.** The results of sensitivity analyses for the mediation analysis.

	<b>Item</b>	<b>ACME/TIE</b>	<b>ADE/TDE</b>	<b>Mediation proportion (%)</b>
4-year	Vitamin D	0.003 (0.002, 0.004)	0.014 (0.001, 0.029)	16.60
	Calcium	0.0007 (0.0005, 0.0009)	0.012 (0.005, 0.027)	5.64
	Brain age acceleration	0.004 (0.0018, 0.006)	0.037 (0.005, 0.060)	9.60
	Combined	0.006 (0.005, 0.008)	0.012 (0.010, 0.015)	33.3
5-year	Vitamin D	0.004 (0.003, 0.005)	0.021 (0.017, 0.025)	16.16
	Calcium	0.0005 (0.0003, 0.0006)	0.010 (0.002, 0.020)	5.15
	Brain age acceleration	0.002 (0.0014, 0.004)	0.011 (0.003, 0.018)	11.0
	Combined	0.005 (0.004, 0.006)	0.018 (0.016, 0.019)	21.7



**Figure S1.** The flowchart of participants selection in the study.



**Figure S2.** The directed acyclic graph for covariate selection.

## References

1. Imbens, G.W. The role of the propensity score in estimating dose-response functions. *Biometrika* **2000**, *87*, 706–710.
2. Hirano, K.; Imbens, G.W. The propensity score with continuous treatments. *Appl. Bayesian Model. Causal Inference Incomplete-Data Perspect.* **2004**, *226164*, 73–84.
3. Austin, P.C. Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat. Methods Med. Res.* **2019**, *28*, 1365–1377.
4. Gao, Q.; Zhang, Y.; Liang, J.; et al. High-dimensional generalized propensity score with application to omics data. *Brief. Bioinform.* **2021**, *22*, bbab331.
5. Wu, X.; Braun, D.; Schwartz, J.; et al. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci. Adv.* **2020**, *6*, eaba5692.
6. Wang, Y.; Lee, M.; Liu, P.; et al. Doubly Robust Additive Hazards Models to Estimate Effects of a Continuous Exposure on Survival. *Epidemiology* **2017**, *28*, 771–779.