Supplementary Online Content

Lourida I, Hannon E, Littlejohns TJ, et al. Association of lifestyle and genetic risk with incidence of dementia [published online July 14, 2019]. *JAMA*. doi:10.1001/jama.2019.9879

This supplementary material has been provided by the authors to give readers additional information about their work.

Supplement 1

eMethods

Polygenic risk score

UK Biobank genotype data were downloaded from the European Genome-phenome Archive. Alzheimer disease (AD) GWAS statistics based on individuals of European ancestry from Lambert and colleagues[1] were downloaded from http://web.pasteurlille.fr/en/recherche/u744/igap/_igap_download.php. For risk score calculation we followed a standard pipeline for calculating polygenic risk scores[2,3] which requires the use of separate training and testing datasets as described elsewhere.[3.] We used the Lambert and colleagues[1] stage 1 GWAS results as the training dataset from which single nucleotide polymorphisms (SNPs) were selected and the effects on the outcome were estimated. As the test dataset, which contains the genotypes of the individuals for whom risk scores are to be calculated, we used the UK Biobank. To select the SNPs which would be included in the calculation of the polygenic risk score, we first filtered the Lambert and colleagues[1] GWAS results to exclude variants with minor allele frequency < 1%. We then performed "intelligent" SNP pruning, known as clumping, to retain the most significant, independently associated variant present in the UK Biobank data per linkage disequilibrium block, removing additionally associated SNPs in that linkage disequilibrium block that tag the same underlying causal variant. This clumping procedure was performed using the --clump command in PLINK[4] with the genotype data from UK Biobank and the following parameters: --clump-p1 1 --clump-p2 1 --clump-r2 0.2 --clump-kb 1000. As the final step of variant selection, we took all SNPs with $P < 0.5$ from the clumped results to form the basis of the polygenic risk score; this threshold was chosen as previous work has shown it contains the optimal set of variants for predicting AD.[2] In total 249,273 independent variants including the *APOE* region (defined as between 44,400KB-46,500KB on chromosome 19) were used to calculate the polygenic risk score. While this set is likely to include a number of SNPs which are not associated with AD, this is outweighed by the number of SNPs included that at present do not have sufficient statistical evidence to meet the criteria for being genome-wide significant (i.e. P-value < $5x10^{-8}$) but will be associated in future larger studies.

For each individual in the UK Biobank, the number of associated alleles weighted by the log of the odds ratio reported by Lambert and colleagues[1] was counted and summed across all SNPs, implemented with the --score command in PLINK.[4] The polygenic risk scores were then z-standardized, divided into quintiles and categorized as low (lowest quintile), intermediate (quintiles 2 to 4) and high (highest quintile).

Genetic principal components in the UK Biobank were calculated across all genetic variants that remained after clumping the AD GWAS results as this provided an independent set of SNPs, using the GCTA software.[5] Considering only the samples self-reported as British, Irish or other white background (measured at baseline with an amalgam of sequential questions from the UK Biobank Touchscreen Questionnaire), filtering thresholds were defined for the first two genetic principal components as two standard deviations from the mean and excluded samples outside of this area to create a genetically homogeneous sample.

References

1. Lambert J-C, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013;45(12):1452.

2. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 2015;138(Pt 12):3673-3684.

3. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;460(7256):748-52.

4. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81(3):559-75.

5. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88(1):76-82.

eTable 1. Healthy lifestyle factor definitions

Healthy lifestyle factor	Source and definition	Self-reported UK Biobank field code
No current smoking	UK Biobank Touchscreen questionnaire at baseline; smoking status defined as current, previous, never smoker	20116
Regular physical activity	UK Biobank Touchscreen questionnaire at baseline; ≥150 minutes moderate activity per week OR ≥ 75 minutes vigorous activity per week OR equivalent combination OR moderate physical activity at least 5 days a week or vigorous activity once a week	884, 894, 904, 914
Healthy diet	UK Biobank Food Frequency Questionnaire at baseline; At least 4 of the following 7 food groups: 1. Fruits: ≥ 3 servings/day 2. Vegetables: ≥ 3 servings/day 3. Fish: ≥2 servings/week 4. Processed meats: ≤ 1 serving/week 5. Unprocessed red meats: ≤ 1.5 servings/week 6. Whole grains: ≥ 3servings/day 7. Refined grains: ≤1.5servings/day	1309, 1319, 1289, 1299, 1329, 1339, 1349, 1369, 1379, 1389, 1438, 1448, 1458, 1468
Moderate alcohol consumption	UK Biobank Touchscreen questionnaire at baseline; US Dietary guidelines for Americans 2015-2020 of up to 1 drink/day for women and up to 2 drinks/day for men. To calculate drink-equivalents as per guidelines, multiply the volume in ounces by the alcohol content in percent and divide by 0.6 ounces of alcohol per drink-equivalent; then convert to grams: 1 drink-equivalent described as containing 14g of pure alcohol. 125ml wine=0.85 drink-equivalents, 4% ABV pint beer = 1.28 drink-equivalents, 25ml spirits=0.57 drink-equivalents, 50ml fortified wine= 0.56 drink-equivalents, Moderate consumption: women: >0 and ≤14g/day men: >0 and ≤28g/day	20117, 1558, 1568, 1578, 1588, 1598, 1608, 5364, 4407, 4418, 4429, 4440, 4451, 4462

Abbreviations: ABV, alcohol by volume.

eTable 2. Codes used in the UK Biobank study to identify dementia cases

ICD-9	ICD-10	Self-reported UK Biobank field code
AD: 331.0 VaD: 290.4 FTD: 331.1 Other codes for all-cause dementia: 290.2, 290.3, 291.2, 294.1, 331.2, 331.5	AD: F00, F00.0, F00.1, F00.2, F00.9, G30, G30.0, G30.1, G30.8, G30.9 VaD: F01, F01.0, F01.1, F01.2, F01.3, F01.8, F01.9, I67.3 FTD: F02.0, G31.0, Other codes for all-cause dementia: A81.0, F02, F02.1, F02.2, F02.3, F02.4, F02.8, F03, F05.1, F10.6, G31.1, G31.8	1263

Abbreviations: ICD, International Classification of Diseases; AD, Alzheimer disease, VaD, vascular dementia, FTD, frontotemporal dementia.

eTable 3. Associations between the polygenic risk score and individual lifestyle factors

	OR (95% CI)	*P*-value
No smoking		
Polygenic risk score	1.00 (0.98-1.01)	0.65
Regular physical activity		
Polygenic risk score	1.01 (1.00-1.02)	0.02
Healthy diet		
Polygenic risk score	1.00 (0.99-1.01)	0.37
Moderate alcohol consumptio		
Polygenic risk score	0.99 (0.98-1.00)	0.26

Note: Adjusted for age, sex, education, socioeconomic status, relatedness, number of alleles included in the polygenic risk score, first 20 principal components of ancestry and the remaining lifestyle factors.
Abbreviations: OR, odds ratio; CI, confidence interval.

eTable 4. Risk of incident dementia according to genetic risk quintile

Genetic risk quintiles	HR (95% CI)	*P*-value	*P*-trend
Quintile 1 (lowest)	Reference		
Quintile 2	1.29 (1.09-1.52)	0.03	
Quintile 3	1.40 (1.19-1.65)	<0.001	<0.001
Quintile 4	1.43 (1.22-1.68)	<0.001	
Quintile 5 (highest)	1.91 (1.64-2.23)	<0.001	

Note: Adjusted for age, sex, education, socioeconomic status, relatedness, number of alleles included in the polygenic risk score, and first 20 principal components of ancestry.
Abbreviations: HR, hazard ratio; CI, confidence interval.

eTable 5. Risk of incident dementia according to healthy lifestyle score

Healthy lifestyle score	HR (95% CI)	*P*-value	*P*-trend
0 healthy lifestyle factor	Reference		
1 healthy lifestyle factor	0.84 (0.54-1.29)	0.42	
2 healthy lifestyle factors	0.70 (0.46-1.05)	0.09	0.002
3 healthy lifestyle factors	0.65 (0.43-0.98)	0.04	
4 healthy lifestyle factors	0.64 (0.43-0.97)	0.04	

Note: Adjusted for age, sex, education, socioeconomic status, relatedness and first 20 principal components of ancestry.
Abbreviations: HR, hazard ratio; CI, confidence interval.

eTable 6. Total participants and incident dementia cases in each genetic and lifestyle category

	Low genetic risk			Intermediate genetic risk			High genetic risk		
Lifestyle	Favorable	Intermediate	Unfavorable	Favorable	Intermediate	Unfavorable	Favorable	Intermediate	Unfavorable
Participants, Total No. [a]	26,856	9,114	3,165	80,290	27,703	9,603	26,407	9,380	3,373
No. Dementia Cases[b] **/Person-years**	151/ 211,986	57/ 72,342	29/ 24,460	635/ 633,405	280/ 219,777	99/ 74,005	298/ 208,769	111/ 74,652	60/ 26,039
Absolute risk, % (95% CI)	0.56 (0.48-0.66)	0.63 (0.48-0.81)	0.92 (0.64-1.31)	0.79 (0.73-0.85)	1.01 (0.90-1.14)	1.03 (0.85-1.25)	1.13 (1.01-1.26)	1.18 (0.98-1.42)	1.78 (1.38-2.28)
Incidence rates per 1,000 person-years (95% CI)	0.71 (0.61-0.84)	0.79 (0.61-1.02)	1.19 (0.82-1.71)	1.00 (0.93-1.08)	1.27 (1.13-1.43)	1.34 (1.10-1.63)	1.43 (1.27-1.60)	1.49 (1.23-1.79)	2.30 (1.79-2.97)

[a] Total No. varies among imputations; low genetic risk: 26,856-26,989 for favorable lifestyle, 9,114-9,243 for intermediate lifestyle and 3,165-3,182 for unfavorable lifestyle; intermediate genetic risk: 80,290-80,515 for favorable lifestyle, 27,703-27,918 for intermediate lifestyle and 9,603-9,634 for unfavorable lifestyle; high genetic risk: 26,407-26,517 for favorable lifestyle, 9,380-9,494 for intermediate lifestyle and 3,373-3,389 for unfavorable lifestyle.

[b] No. of dementia cases varies among imputations; low genetic risk: 151-160 for favorable lifestyle, 57-67 for intermediate lifestyle and 29-30 for unfavorable lifestyle; intermediate genetic risk: 635-659 for favorable lifestyle, 280-304 for intermediate lifestyle and 99-102 for unfavorable lifestyle; high genetic risk: 298-313 for favorable lifestyle, 111-126 for intermediate lifestyle and 60-61 for unfavorable lifestyle.

Abbreviations: CI, confidence interval.

eTable 7. Risk of incident dementia according to genetic and lifestyle risk based on an unweighted and weighted lifestyle score in imputed and unimputed data

	Unweighted lifestyle score (imputed data) n = 196,383		**Unweighted lifestyle score (unimputed data) n = 162,268**		**Weighted lifestyle score (unimputed data) n = 162,268**	
Subgroup	**Hazard Ratio (95% CI)**	***P*-value**	**Hazard Ratio (95% CI)**	***P*-value**	**Hazard Ratio (95% CI)**	***P*-value**
Low genetic risk group						
Favorable lifestyle	1.00	Reference	1.00	Reference	1.00	Reference
Intermediate lifestyle	1.02 (0.74-1.41)	0.89	1.19 (0.87-1.63)	0.27	1.30 (0.95-1.80)	0.11
Unfavorable lifestyle	1.20 (0.79-1.83)	0.39	1.41 (0.91-2.17)	0.12	1.69 (1.09-2.62)	0.02
Intermediate genetic risk						
Favorable lifestyle	1.32 (1.09-1.58)	0.004	1.31 (1.06-1.61)	0.01	1.34 (1.10-1.64)	0.004
Intermediate lifestyle	1.51 (1.23-1.85)	<0.001	1.62 (1.29-2.02)	<0.001	1.83 (1.47-2.29)	<0.001
Unfavorable lifestyle	1.73 (1.35-2.23)	<0.001	1.93 (1.47-2.52)	<0.001	1.87 (1.41-2.49)	<0.001
High genetic risk						
Favorable lifestyle	1.92 (1.56-2.35)	<0.001	1.88 (1.49-2.36)	<0.001	1.90 (1.52-2.37)	<0.001
Intermediate lifestyle	1.84 (1.43-2.35)	<0.001	1.85 (1.41-2.43)	<0.001	2.05 (1.55-2.70)	<0.001
Unfavorable lifestyle	2.62 (1.91-3.58)	<0.001	2.94 (2.13-4.06)	<0.001	3.09 (2.21-4.34)	<0.001

Note: Adjusted for age, sex, education, socioeconomic status, relatedness, number of alleles included in the polygenic risk score, and first 20 principal components of ancestry.
Abbreviations: HR, hazard ratio; CI, confidence interval.

eTable 8. Risk of incident dementia according to genetic and lifestyle risk with additional adjustment for depressive symptoms, stroke history, after excluding related participants and with higher cut-point for moderate alcohol consumption

	Depressive symptoms[a] n = 196,383		**History of stroke[b] n = 196,383**		**Related participants excluded[c] n = 132,945**		**Moderate alcohol consumption >0 and <38g/day[d] n = 196,383**	
	HR (95% CI)	***P*-value**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**
Low genetic risk								
Favorable lifestyle	Reference				Reference		Reference	
Intermediate lifestyle	1.09 (0.79-1.50)	0.60	1.10 (0.80-1.51)	0.57	1.03 (0.69-1.54)	0.90	1.36 (1.01-1.84)	0.04
Unfavorable lifestyle	1.47 (0.99-2.19)	0.06	1.47 (0.98-2.18)	0.06	1.77 (1.12--2.79)	0.01	1.91 (1.22-2.99)	0.005
Intermediate genetic risk								
Favorable lifestyle	1.36 (1.14-1.63)	0.001	1.36 (1.14-1.63)	0.001	1.35 (1.09--1.68)	<0.001	1.39 (1.16-1.66)	<0.001
Intermediate lifestyle	1.67 (1.37-2.05)	<0.001	1.68 (1.37-2.05)	<0.001	1.62 (1.26-2.08)	<0.001	1.95 (1.59-2.38)	<0.001
Unfavorable lifestyle	1.63 (1.26-2.11)	<0.001	1.64 (1.27-2.12)	<0.001	1.67 (1.22-2.28)	0.001	2.20 (1.66-2.91)	<0.001
High genetic risk								
Favorable lifestyle	1.95 (1.60-2.38)	<0.001	1.96 (1.61-2.39)	<0.001	2.01 (1.58-2.55)	<0.001	1.94 (1.59-2.35)	<0.001
Intermediate lifestyle	1.99 (1.55-2.55)	<0.001	2.00 (1.56-2.57)	<0.001	2.08 (1.54-2.81)	<0.001	2.63 (2.07-3.33)	<0.001
Unfavorable lifestyle	2.71 (2.00-3.67)	<0.001	2.74 (2.03-3.71)	<0.001	2.59 (1.77-3.79)	<0.001	3.21 (2.23-4.61)	<0.001

Abbreviations: HR, hazard ratio; CI, confidence interval.

[a] Also adjusted for age, sex, education, socioeconomic status, depressive symptoms, relatedness, number of alleles included in the polygenic risk score, and first 20 principal components of ancestry.

[b] Also adjusted for age, sex, education, socioeconomic status, relatedness, number of alleles included in the polygenic risk score, and first 20 principal components of ancestry.

[c] Adjusted for age, sex, education, socioeconomic status, number of alleles included in the polygenic risk score, and first 20 principal components of ancestry.

[d] Adjusted for age, sex, education, socioeconomic status, relatedness, number of alleles included in the polygenic risk score, and first 20 principal components of ancestry.

eTable 9. Risk of incident dementia according to genetic and lifestyle risk stratified by sociodemographic variables

	Aged 60-64 n = 109,769		Aged 65+ n = 86,614		Women n = 103,443		Men n = 92,940		Low education n = 114,908[a]		High education n = 81,368[b]	
Subgroup	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**
Low genetic risk												
Favorable lifestyle	Reference		Reference		Reference		Reference		Reference		Reference	
Intermediate lifestyle	0.73 (0.40-1.33)	0.30	1.31 (0.90-1.90)	0.15	1.48 (0.93-2.35)	0.10	0.88 (0.57-1.36)	0.56	1.12 (0.77-2.65)	0.06	1.05 (0.58-1.89)	0.87
Unfavorable lifestyle	1.13 (0.55-2.29)	0.74	1.74 (1.07-2.81)	0.03	2.33 (1.30-4.20)	0.005	1.11 (0.64-1.91)	0.72	1.42 (0.89-2.27)	0.14	1.76 (0.81-3.86)	0.16
Intermediate genetic risk												
Favorable lifestyle	1.23 (0.90-1.66)	0.19	1.43 (1.15-1.79)	0.001	1.71 (1.29-2.26)	<0.001	1.15 (0.91-1.45)	0.25	1.29 (1.03-1.61)	0.02	1.51 (1.11-2.05)	0.009
Intermediate lifestyle	1.42 (1.00-2.01)	0.05	1.85 (1.45-2.37)	<0.001	1.87 (1.35-2.59)	<0.001	1.60 (1.24-2.07)	<0.001	1.66 (1.30-2.12)	<0.001	1.76 (1.22-2.53)	0.002
Unfavorable lifestyle	1.53 (1.00-2.33)	0.05	1.76 (1.28-2.43)	0.001	1.93 (1.26-2.96)	0.002	1.53 (1.11-2.11)	0.009	1.55 (1.14-2.11)	0.005	2.13 (1.33-3.40)	0.002
High genetic risk												
Favorable lifestyle	1.51 (1.06-2.14)	0.02	2.19 (1.72-2.78)	<0.001	2.23 (1.63-3.04)	<0.001	1.78 (1.38-2.30)	<0.001	1.96 (1.54-2.50)	<0.001	1.90 (1.34-2.71)	<0.001
Intermediate lifestyle	1.59 (1.02-2.49)	0.04	2.24 (1.65-3.04)	<0.001	2.32 (1.58-3.42)	<0.001	1.82 (1.31-2.52)	<0.001	1.89 (1.40-2.55)	<0.001	2.30 (1.46-3.61)	<0.001
Unfavorable lifestyle	2.66 (1.62-4.36)	<0.001	2.88 (1.97-4.22)	<0.001	3.99 (2.53-6.31)	<0.001	2.18 (1.45-3.27)	<0.001	2.59 (1.82-3.70)	<0.001	3.67 (2.05-6.56)	<0.001

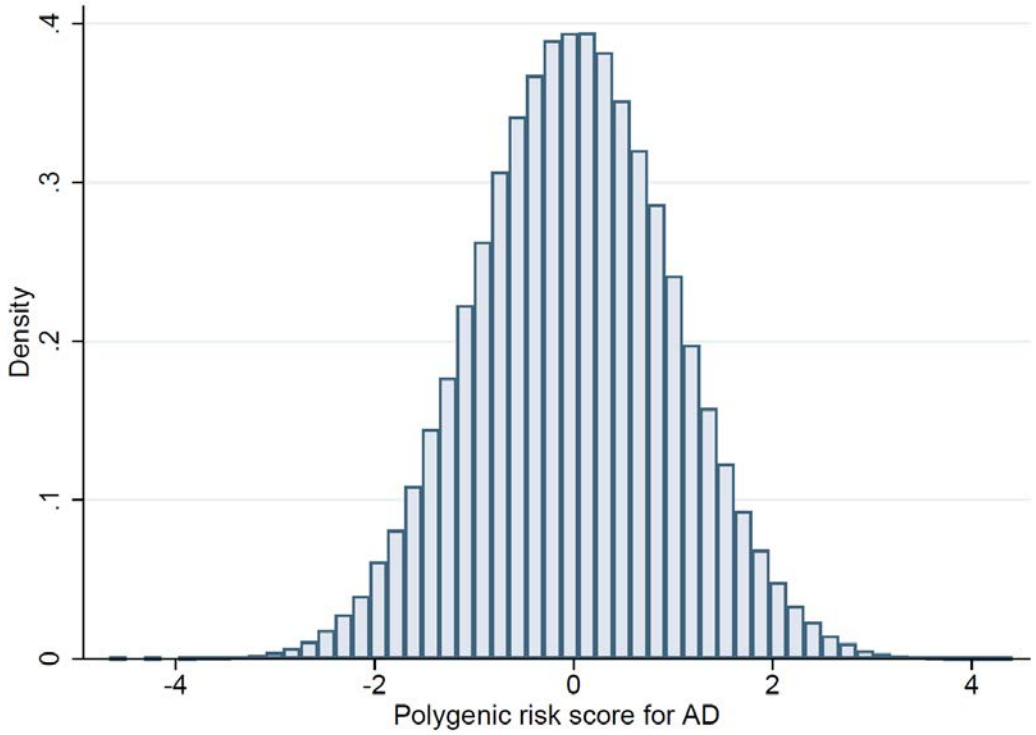
Note: Adjusted for age, sex, education, socioeconomic status, relatedness, number of alleles included in the polygenic risk score, and first 20 principal components of ancestry unless stratified for one of these variables.

Abbreviations: HR, hazard ratio; CI, confidence interval.

[a] Sample sizes vary between 114,908 and 115,015 across imputations.

[b] Sample sizes vary between 81,368 and 81,475 across imputations.

eFigure 1. Distribution of the polygenic risk score for Alzheimer disease



Abbreviations: AD, Alzheimer disease.