Brief communication: The proportion of Alzheimer's disease attributable to apolipoprotein E

Dylan M. Williams PhD[1], Neil M. Davies PhD[2-4], Emma L. Anderson PhD[2, 5, 6]

[1] MRC Unit for Lifelong Health & Ageing at UCL, University College London, London, UK

[2] Division of Psychiatry, University College London, Maple House, 149 Tottenham Court Rd, London W1T 7NF

[3] K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Norway.

[4] Department of Statistical Science, University College London, London WC1E 6BT, UK

[5] Medical Research Council Integrative Epidemiology Unit at the University of Bristol, BS8 2BN, United Kingdom.

[6] Bristol Medical School, Population Health Sciences, University of Bristol, UK

Corresponding author:

Dylan M. Williams; Principal Research Fellow, MRC Unit for Lifelong Health and Ageing at UCL, University College London, Floor 5, 1-19 Torrington Place, London, WC1E 7HB, UK; Tel. +44(0) 20 7670 5713; dylan.williams@ucl.ac.uk

Abstract

The extent to which genetic variation at the *APOE* locus explains the burden of late-onset Alzheimer's disease (AD) is poorly understood. We provide new estimates of the proportions of AD and all-cause dementia attributable to carriage of ε3 and/or ε4 alleles of *APOE*, using data from 171,128 participants of the UK Biobank cohort study. AD and all-cause dementia were ascertained at baseline and during follow-up via linked electronic health records for up to 16.8 years. We estimate that 74.5% (95% CI: 38.6, 89.4) of AD and 39.5% (95% CI: 10.7, 59.1%) of all-cause dementia burden is attributable to ε3 and ε4 carriage. Thus, differences in the molecular physiology of apolipoprotein E cause most AD and a large fraction of all dementia. Research into this pathway should be prioritised to facilitate dementia prevention.

Main text

The extent to which genetic variation at the *APOE* locus explains the burden of late-onset Alzheimer's disease (AD) is poorly understood. Three major apolipoprotein E (apoE) isoforms exist: ε2, ε3 and ε4. Relative to carriage of ε3 – the most common allele with about 95% prevalence worldwide[1] – AD risk is higher with ε4 carriage (~28% prevalence) and lower with ε2 carriage (~14% prevalence). The proportion of AD cases attributable to the detrimental ε4 allele has been estimated in many settings, with population attributable fractions (PAFs) for this burden ranging considerably, up to approximately 50%.[2-4] However, these estimates do not encapsulate the proportion of AD cases attributable to ε3, which is commonly misperceived as neutral for AD risk, even though ε3 substantially increases risk of AD relative to ε2 carriage.[4,5] Previous PAF estimates have also likely been deflated due to biases in study design – ascertaining AD/dementia through clinical follow-ups alone (prone to attrition), the elimination of prevalent cases, and/or limited follow-up time in incidence studies.[2] PAFs have also been biased downwards when calculated erroneously -- for instance, when using allele frequency (an allele's proportion among *all alleles* in a sample) instead of genotype frequency (i.e. the proportion of individuals in a sample with a given genotype) as the prevalence of the exposure in estimates.[4] Establishing the full extent of the AD burden that arises due to common differences in apoE is important, because it would indicate the proportion of cases that could be prevented by intervening on this single molecular pathway.

Using data from 171,128 participants of the UK Biobank cohort study (UKB) with long-term follow-up via electronic health records, we provide new estimates of the proportions of AD and all-cause dementia attributable to ε3 *and* ε4 carriage – i.e. the fraction of cases due to the combined impact of alleles inherited by the vast majority of individuals. We conducted a cohort study among participants aged 60 to 72 years at baseline in 2006-2010,[6] where AD and all-cause dementia cases were identified from self-report at baseline and follow-up via linked electronic health and death records available up to December 2022

(minimum/maximum follow-up: 12.2 / 16.8 years). We coded *APOE* genotypes as one exposure representing ε3 and/or ε4 carriage, relative to ε2 homozygotes. Individuals with an ε3/ε3 genotype are typically used as the reference group in analyses of *APOE* because ε3/ε3 is the most common genotype. However, to appropriately calculate attributable risk for an exposure with multiple levels, individuals with the lowest risk should be set as the reference group[7] – individuals of ε2/ε2 genotype in this instance. This allows the disease burden attributable to exposure to all risk-increasing genotypes to be calculated. Thus, we modelled the full spectrum of risk of AD and all-cause dementia encompassed by combinations of ε3 and ε4 carriage, relative to the lowest risk group (ε2 homozygotes).

Among the analytical sample, 99.4% had ε3 and/or ε4 carriage (Supplemental Table 1). The odds ratio for AD due to ε3 and/or ε4 carriage was 3.93 (95% CI: 1.63, 9.48; Table 1). The equivalent for all-cause dementia was 1.66 (1.12, 2.45). Together, the two *APOE* risk alleles had PAFs of 74.5% (38.6, 89.4%) and 39.5% (10.7, 59.1%) for AD and all-cause dementia, respectively. In secondary analyses using ascertainment of AD and dementia without diagnoses from primary care records (only available for 45% of the cohort; online methods), the PAFs were slightly lower for AD and slightly higher for all-cause dementia (supplemental table 2).

To put these findings into context, we quantified the proportions of AD attributable to the next nine strongest genetic loci after *APOE*, as identified by large genome-wide association studies of late-onset AD (Figure 1).[4] To compare these magnitudes to effects for another chronic disease, we also calculated PAFs for the proportions of coronary artery disease (CAD) attributable to the disease's ten strongest genetic risk loci. No PAF for any other locus for either condition exceeded 22%. The preponderance of disease burden attributable to genetic variation at a single locus is exceptional among common, complex chronic diseases.

We then extended calculations to estimate the separate contributions of ε3 and ε4 to the overall burden of AD attributable to these two alleles, on the basis of genotype-specific PAFs (table 1). For ε4, this meant estimating disease burden attributable to genotypes ε2/ε4 and ε4/ε4 and a share of the risk among ε3/ε4 carriers due to ε4 specifically – equating to a PAF of 45.5% for AD for ε4 carriage. The remainder of the PAF for AD (29.5%) was attributable to ε3 carriage specifically, due to ε2/ε3 and ε3/ε3 carriage and ε3's contribution to risk among ε3/ε4 individuals.

In summary, our findings indicate that if interventions could obviate entirely the detrimental effects of ε3 and ε4 carriage in a population akin to the UKB sample, we could expect to prevent approximately three-quarters of AD cases. Put differently, if all individuals inherited an ε2/ε2 genotype, most AD would not occur. Such a magnitude of attributable risk has been suggested previously (ε3 and ε4 carriage perhaps accounting for 95% of AD[8]) but not demonstrated directly before. Reasons for this likely include the necessity for very large analytical samples to use rare ε2 homozygotes as the reference group in analyses, and because there is typically little recognition among dementia researchers that ε3 genotypes should also be considered as risk-increasing for AD. Nonetheless, we estimate that the ε3 allele alone could be responsible for almost a third or more of AD, as a result of ε3 conferring considerable risk to a large fraction of the population. It should be noted that PAFs are distinct from heritability, with the heritability of AD estimated to range up to 79%.[9] However, heritability analyses are not informative for assessing disease burden attributable to specific causes and PAFs are more appropriate and intuitive for this purpose.[10]

Limitations of this research include incomplete ascertainment of AD and all-cause dementia cases due to limited record linkage in UKB (which does not yet fully extend to primary care or mental health service records)[11] and no follow-up with cognitive assessments of the whole surviving cohort. Measurement of outcome lifetime risks was incomplete – the youngest participants in our sample were aged 73 years at the end of current follow-up. Recruitment

into UKB was not representative of the general UK population. It may have been affected by selection effects,[12] perhaps including effects from *APOE* genotypes contributing to cardiovascular and other morbidity and mortality before age 60 years.[13] Some AD cases may have been misclassified and had dementia of other aetiologies; *APOE* associations with AD are stronger in samples where cases have been neuropathologically confirmed.[14] However, these biases will likely lead to underestimated, rather than inflated, PAFs. Attributable fractions also assume that the exposure of interest is a cause of the disease being investigated, and not based on biased estimates of risk.[7] Due to the properties of genetic inheritance, risk estimates for genetic variants such as the *APOE* alleles are not subject to reverse causation and are unlikely to be affected by confounding.[15] The effects of variation in *APOE* on AD and all-cause dementia risk are also highly unlikely to be due to nearby co-inherited genetic variation (in linkage disequilibrium with *APOE* variation), rather than the *APOE* variants per se.[16] Hence, PAFs for these variants provide more robust estimates of disease burden attributable to the variation in question than PAFs estimated for environmental factors, for instance.[17]

Therefore, given that most AD appears to be caused by differences in the molecular physiology of apoE, our findings and others[18] should motivate proportionate attention and funding for research into the mechanisms linking apoE with AD. These should include efforts to understand the distinct functional properties of the ε3 isoform that confer AD risk – relative to properties of the ε2 isoform and other protective variants [18,19] – and not only further research to elucidate and mitigate ε4's effects. We note that it is often incorrectly assumed that genetic risk is unmodifiable.[20] With the advent of gene editing and silencing, genetic risk may now be directly modifiable; thus, editing *APOE* alleles or affecting the gene's expression in relevant cells at the most pertinent life stage(s) could potentially prevent most cases of AD. Moreover, genetic findings point to molecular physiology to target by other means. For instance, mutations in the gene *PCSK9* that cause familial hypercholesterolaemia led to the development of proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors.[21] Many

strategies to target apoE exist, including immunotherapy and small molecule structural correctors.[22,23] However, only one therapy targeting apoE (LX1001) is currently being trialled in humans – representing less than 1% of potential therapies for AD in registered trials.[24] To reiterate, findings such as ours should prompt a rebalancing of therapeutic development for AD (as well as basic research) towards apoE. Prioritising direct research into apoE should not preclude research into broader genetic or environmental factors that could be mediating or modifying the effects of apoE on AD or investigations into factors that may be distinct causes of these outcomes independently of apoE (both scenarios include research addressing cerebral amyloidosis and tauopathy). Nonetheless, establishing precisely how, when and in which cell types apoE influences AD risk – and how its deleterious effects can be mitigated – is paramount to AD prevention and treatment.

Methods

Study design

UKB is a multi-centre cohort study that recruited approximately 502,000 participants aged 39–73 years at assessment sites in England, Scotland, and Wales between 2006 and 2010.[6] Here, we studied data from a sub-set that were aged ≥60 years at baseline with genotypic data, after exclusions for failing sample-level genetic quality control (genetic/phenotypic sex mismatches, excess heterozygosity, aneuploidy) and the random removal of one individual from related pairs. *APOE* ε2/ε3/ε4 alleles were coded from genotyped or hard-called imputed microarray data for single nucleotide polymorphisms rs7412 and rs429358.[6] All-cause dementia was identified using the cohort's algorithmically defined outcomes from a combination of self-report at baseline and follow-up via linked electronic health and death records available up to December 2022 (minimum/maximum follow-up: 12.2 / 16.8 years). Alzheimer's disease, more specifically, was ascertained only via record linkage and not by baseline self-reporting, due to wording of the survey item at the baseline assessment enquiring about history of 'dementia or Alzheimer's disease or cognitive impairment' non-specifically. In primary analyses, ascertainment of both outcomes was expanded to include

identification of diagnostic codes from primary care records for the ~45% of the cohort for which linkage to records from general practices has been arranged (code lists in supplemental table 3; linkage is being sought for the remainder of the cohort). In secondary (sensitivity) analyses, we limited AD and dementia ascertainment to the use of self-report and secondary care / death record data available for the entire cohort.

Statistical analysis

Risk of AD and all-cause dementia were modelled using multivariable logistic regression, adjusting for age at baseline, sex, ethnicity (entered as a binary variable for white / other in models due to small numbers of ethnic minorities being present in each outcome group in sex-specific and individual *APOE* genotype analyses), the first ten genetic principal components supplied by the UK Biobank team,[6] and genotyping array. Logistic regression was adopted for analyses, rather than survival analysis, because of the sample's mix of prevalent and incident cases. The prevalence of the exposure (genotype frequency) in the full sample and computed odds ratios for AD and all-cause dementia were used to calculate population attributable fractions (PAF)[7]:

$$AF\ (population)\ = \frac{Prevalence_{exposure}\ (OR-1)}{1+Prevalence_{exposure}\ (OR-1)} \qquad (1)$$

95% confidence intervals for PAFs were derived using the lower and upper confidence intervals for the odds ratios. PAFs and their confidence intervals were converted from fractions to percentages. Given the low prevalence of the outcomes in the sample (and particularly in the reference group), odds ratios were regarded as equivalent to risk ratios for PAF calculations. We also stratified models by sex to produce PAFs for each outcome in females and males separately.

To calculate PAFs for AD attributable to other genomic loci besides *APOE*, we used data for the top nine loci beyond *APOE* from one of the largest genome-wide association studies (GWAS) of neuropathologically confirmed AD.[4] For the risk-increasing allele of each GWAS

hit, we first identified allele frequency *p* among individuals of European ancestry in the 1000 Genomes project, phase 3.[1] We then calculated genotype frequencies for homozygous carriers of the risk allele (p^2) and heterozygous carriers of the risk allele ($2 \times p \times (1-p)$). On the basis that the GWAS modelled risk assuming additive effects of variants, we estimated overall PAFs for each variant as a sum of a PAF for homozygous carriers and a PAF for heterozygous carriers. The genotype frequencies of homozygotes and heterozygotes were entered as prevalence of the exposure in equation (1). For homozygotes, odds ratios were recalculated as exp(2 × log-odds) reported by the GWAS; for heterozygotes, odds ratios were based on the reported log-odds. We note that our PAF calculations for AD risk loci differ from the authors' own calculations which are presented in supplemental table 6 of their publication.[4] These appear to have been based on allele frequency among controls in their sample instead of genotype frequencies for exposure to an allele, and the use of a single odds ratio rather than doubling of the odds for the proportion of individuals that would be homozygous for a risk allele – hence, these PAFs are substantially lower than ours. To calculate the equivalent statistics for coronary artery disease (CAD), we used summary statistics from one of the largest CAD GWAS to date.[25] We used the same approach as for AD risk loci to calculate PAFs for the top 50 loci identified from this GWAS, and then ranked these to use the highest 10 PAFs in our figure.

To evaluate the separate contributions of ε3 and ε4 alleles to the overall burden of AD attributable to these two *APOE* alleles, we used a formula for calculating PAFs for multi-level exposures.[26] We applied this to each of the five risk-increasing genotypes (ε2/ε3, ε2/ε4, ε3/ε3, ε3/ε4, ε4/ε4), relative to ε2/ε2 carriage, with an indicator variable for genotype entered into the logistic regression modelling described above. The contributions of ε3 and ε4 were then calculated as the sums of PAFs from individual genotypes including each allele separately (e,g, ε2/ε3 and ε3/ε3 for ε3 carriage) along with the estimated shares of each allele due to the increased risk experienced by ε3/ε4 carriers. The genotype-specific PAF for ε3/ε4 was partitioned into ε4 and ε3 contributions according to the ratio by which ε2/e4 to

ε2/ε3 genotypes increase AD risk, i.e. according to the individual effects of the two alleles on AD risk. This ratio was 2.46:1 for ε4:ε3, meaning that ε4 was estimated to be responsible for 71.1% of the increased risk experienced by ε3/ε4 carriers – and hence 71.1% of the PAF for the ε3/ε4 genotype, with the remainder attributed to ε3. This assumes that any contribution to an interaction between the two alleles in relation to AD risk is proportionate to each allele's effects on AD risk in isolation among ε2/ε4 and ε2/ε3 carriers. We note that the overall PAF for AD due to genotypes containing ε3 and/or ε4 in this analysis (75.0%) differed slightly to the point estimate in our main analysis (74.5%) due to differences in precision when making five individual comparisons rather than the use of one binary exposure.

Ethics

UKB participants had given written informed consent and ethical approval for the study was granted by the North West Haydock Research Ethics Committee of the UK's Health Research Authority.

Missingness

From a starting sample of 217,458 individuals in UK Biobank aged 60 years or older at baseline, 171,128 were included in the analytical sample (78.7%). Sequential exclusions were based on: i) no genotype data available (N=6,242); ii) individuals removed during GWAS quality control steps (N=39,242; primarily due to relatedness among study participants); iii) missing data on self-reported ethnicity (N=841); iv) exclusion of individuals whose genotype data implied that they may have a rare ε3r (also known as ε1) allele (N=5).

We followed STREGA reporting recommendations for genetic association studies.

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 71702. The MRC Unit for Lifelong Health and Ageing at UCL is funded by the Medical Research Council (MC_UU_00019/3). NMD is supported via a Norwegian Research Council Grant number 295989. ELA is supported by a UKRI Future Leaders Fellowship (MR/W011581/1). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Author information

DMW conceived the study, undertook the analyses and drafted the manuscript. All authors contributed to the interpretation of data and the manuscript's content, and approved its final version.

We have no conflicts of interest to disclose.

Data sharing statement

All data used in this research are available to researchers who register with UK Biobank and request access to them as part of an approved project: https://www.ukbiobank.ac.uk/. The data fields and script used in the analyses can be viewed at: *GitHub link to be inserted upon M/S acceptance*.

References

1 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68 (2015).

2 Ashford, J. W. APOE genotype effects on Alzheimer's disease onset and epidemiology. *Journal of Molecular Neuroscience* **23**, 157-165 (2004).

3 Saddiki, H. *et al.* Age and the association between apolipoprotein E genotype and Alzheimer disease: a cerebrospinal fluid biomarker–based case–control study. *PLoS Medicine* **17**, e1003289 (2020).

4 Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45**, 1452-1458 (2013).

5 Farrer, L. A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *JAMA* **278**, 1349-1356 (1997).

6 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

7 Steenland, K. & Armstrong, B. An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology*, 512-519 (2006).

8 Raber, J., Huang, Y. & Ashford, J. W. ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiology of Aging* **25**, 641-650 (2004).

9 Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry* **63**, 168-174 (2006).

10 Björk, J., Andersson, T. & Ahlbom, A. Commonly used estimates of the genetic contribution to disease are subject to the same fallacies as bad luck estimates. *European Journal of Epidemiology* **34**, 987-992 (2019).

11 Wilkinson, T. *et al.* Identifying dementia cases with routinely collected health data: a systematic review. *Alzheimer's & Dementia* **14**, 1038-1051 (2018).

12 Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology* **186**, 1026-1034 (2017). https://doi.org:10.1093/aje/kwx246

13 Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics* **49**, 1385-1391 (2017). https://doi.org:10.1038/ng.3913

14 Reiman, E. M. *et al.* Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. *Nature Communications* **11**, 667 (2020). https://doi.org:10.1038/s41467-019-14279-8

15 Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* **33**, 30-42 (2004).

16 Belloy, M. E., Napolioni, V. & Greicius, M. D. A quarter century of APOE and Alzheimer's disease: progress to date and the path forward. *Neuron* **101**, 820-838 (2019).

17 Livingston, G. *et al.* Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet* **396**, 413-446 (2020).

18 Arboleda-Velasquez, J. F. *et al.* Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. *Nature Medicine* **25**, 1680-1683 (2019).

19 Le Guen, Y. *et al.* Association of rare APOE missense variants V236E and R251G with risk of Alzheimer disease. *JAMA neurology* **79**, 652-663 (2022).

20 Editorial. It's all druggable. *Nature Genetics* **49**, 169-169 (2017). https://doi.org:10.1038/ng.3788

21 Sabatine, M. S. PCSK9 inhibitors: clinical evidence and implementation. *Nature Reviews Cardiology* **16**, 155-165 (2019).

22 Marino, C. *et al.* APOE Christchurch-mimetic therapeutic antibody reduces APOE-mediated toxicity and tau phosphorylation. *Alzheimer's & Dementia* (2023).

23 Wang, C. *et al.* Gain of toxic apolipoprotein E4 effects in human iPSC-derived neurons is ameliorated by a small-molecule structure corrector. *Nature Medicine* **24**, 647-657 (2018).

24 Cummings, J. *et al.* Alzheimer's disease drug development pipeline: 2023. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **9**, e12385 (2023).

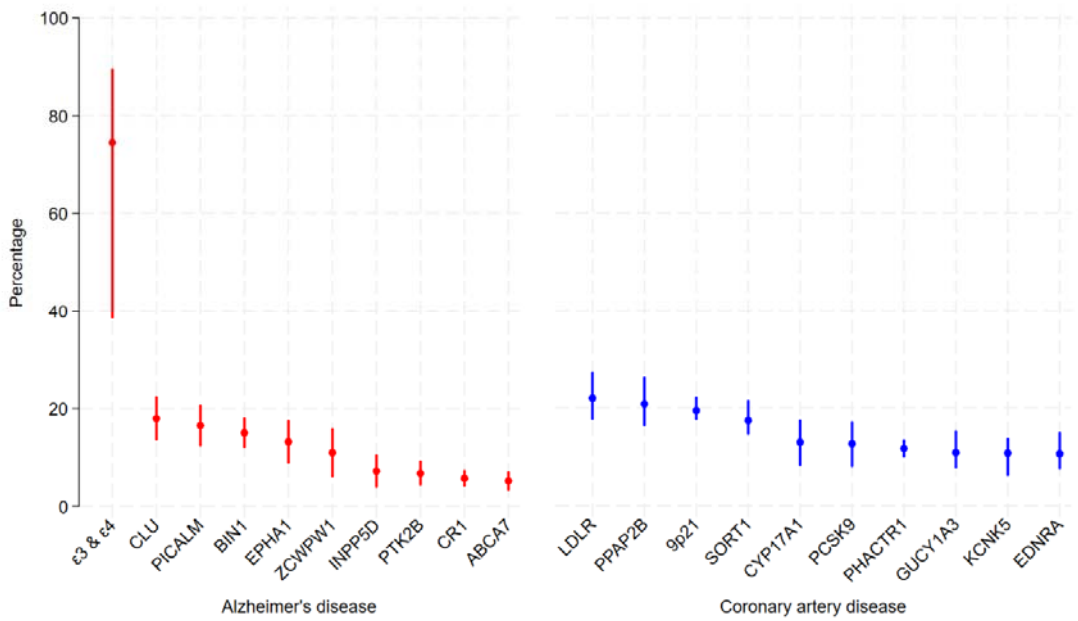
25 The CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* **47**, 1121-1130 (2015).

26 Hanley, J. A heuristic approach to the formulas for population attributable fraction. *Journal of Epidemiology & Community Health* **55**, 508-514 (2001).

Table 1. Odds ratios (OR) and population attributable fractions (PAF) for AD or all-cause dementia by *APOE* genotypes

Exposure	N	AD				All-cause dementia			
		OR	95% CI	P	PAF, % (95%CI)	OR	95% CI	P	PAF, % (95%CI)
1. With ε3 and ε4 grouped									
ε2/ε2	1053	Ref.				Ref.			
ε3 and/or ε4 carriage	170075	3.93	(1.63, 9.48)	0.002	74.5 (38.6, 89.4)	1.66	(1.12, 2.45)	0.01	39.5 (10.7, 59.1)
2. By individual genotype									
ε2/ε2	1053	Ref.				Ref.			
ε2/ε3	21009	1.66	(0.68, 4.05)	0.27	2.0 (0.0, 3.9)	0.91	(0.61, 1.36)	0.65	0.0 (0.0, 1.8)
ε2/ε4	4182	4.09	(1.65, 10.12)	0.002	1.9 (1.0, 2.3)	1.93	(1.27, 2.93)	0.002	1.3 (0.6, 1.9)
ε3/ε3	100892	2.06	(0.85, 4.98)	0.11	15.6 (0.0, 24.2)	1.07	(0.73, 1.59)	0.72	2.6 (0.0, 13.9)
ε3/ε4	39954	8.03	(3.33, 19.38)	<0.001	41.0 (32.7, 44.3)	2.96	(2.00, 4.39)	<0.001	27.0 (20.4, 31.4)
ε4/ε4	4038	25.64	(10.58, 62.16)	<0.001	14.5 (13.6, 14.9)	8.83	(5.92, 13.16)	<0.001	10.9 (10.1, 11.4)

Figure 1. The proportions of Alzheimer's and coronary artery diseases that could be prevented by targeting the pathways encoded by the top ten strongest genetic loci for each condition



Note: attributable fractions for each disease can sum to greater than 100%. This implies that there are interactions between the molecular pathways affected by these genes, as is known to be the case for *LDLR* and *PCSK9* in the determination of circulating low-density lipoprotein cholesterol concentrations and downstream risk of coronary artery disease, for example.

Supplemental table 1. Descriptive characteristics on the sub-set of UK Biobank participants in this analysis (N=171,128)

Characteristic	Mean (SD)	N (%)
Age (years) at baseline	64.1 (2.8)	
Female sex		89,318 (52.1)
Ethnicity		
Asian		2,369 (1.4)
Black		1,448 (0.9)
Chinese		284 (0.2)
Mixed		535 (0.3)
Other [a]		929 (0.5)
White		165,563 (96.8)
APOE genotype [b]		
ε2 carriage		26,244 (15.3)
ε3 carriage		161,855 (94.6)
ε4 carriage		48,174 (28.2)
ε2 homozygous		1,053 (0.6)
ε3 homozygous		100,892 (59.0)
ε4 homozygous		4,038 (2.4)
Alzheimer's disease		3,126 (1.8)
All-cause dementia		6,859 (4.0)

[a] 'Other' refers to anyone who did not select another ethnic group (no more information on ethnic background of these individuals was available)

[b] NB: carriage groups overlap and hence percentages sum to over 100%

Supplemental table 2. Odds ratios and population attributable fractions (PAFs) for AD or all-cause dementia by *APOE* genotypes, omitting the use of primary care records for outcome ascertainment

Exposure	N	AD				All-cause dementia			
		OR	95% CI	P	PAF, % (95%CI)	OR	95% CI	P	PAF, % (95%CI)
1. With ε3 and ε4 grouped									
ε2/ε2	1053	Ref.				Ref.			
ε3 and/or ε4 carriage	171128	3.81	(1.58, 9.18)	0.003	73.6 (36.5, 89.0)	1.74	(1.16, 2.61)	0.008	42.4 (13.6, 61.6)
2. By individual genotype									
ε2/ε2	1053	Ref.				Ref.			
ε2/ε3	21118	1.62	(0.66, 3.96)	0.29	2.0 (0.0, 3.9)	0.96	(0.63, 1.45)	0.84	0.0 (0.0, 2.1)
ε2/ε4	4182	4.03	(1.63, 9.99)	0.003	1.9 (1.0, 2.3)	2.06	(1.33, 3.16)	0.001	1.4 (0.7, 2.0)
ε3/ε3	100892	2.01	(0.83, 4.86)	0.12	15.4 (0.0, 24.4)	1.11	(0.74, 1.67)	0.61	3.7 (0.0, 14.8)
ε3/ε4	39954	7.72	(3.20, 18.63)	<0.001	40.5 (32.0, 44.0)	3.14	(2.08, 4.72)	<0.001	28.0 (21.4, 32.3)
ε4/ε4	4038	24.46	(10.09, 59.30)	<0.001	14.3 (13.4, 14.7)	9.40	(6.21, 14.24)	<0.001	11.1 (10.4, 11.6)

Supplemental table 3: primary care diagnostic codes used to ascertain AD and all-cause dementia

Version	Code	Term	Code included in AD definition
Read V3 (CTV3)	1461.	H/O: dementia	
Read V3 (CTV3)	A410.	Kuru	
Read V3 (CTV3)	A411.	Creutzfeldt-Jakob disease	
Read V3 (CTV3)	E00..	Senile and presenile organic psychot conditions (& dementia)	
Read V3 (CTV3)	E000.	Uncomplicated senile dementia	
Read V3 (CTV3)	E001.	Presenile dementia	
Read V3 (CTV3)	E0010	Uncomplicated presenile dementia	
Read V3 (CTV3)	E0011	Presenile dementia with delirium	
Read V3 (CTV3)	E0012	Presenile dementia with paranoia	
Read V3 (CTV3)	E0013	Presenile dementia with depression	
Read V3 (CTV3)	E001z	Presenile dementia NOS	
Read V3 (CTV3)	E002.	Senile dementia with depressive or paranoid features	
Read V3 (CTV3)	E0020	Senile dementia with paranoia	
Read V3 (CTV3)	E0021	Senile dementia with depression	
Read V3 (CTV3)	E002z	Senile dementia with depressive or paranoid features NOS	
Read V3 (CTV3)	E003.	Senile dementia with delirium	
Read V3 (CTV3)	E004.	Arteriosclerotic dementia (including [multi infarct dement])	
Read V3 (CTV3)	E0040	Uncomplicated arteriosclerotic dementia	
Read V3 (CTV3)	E0041	Arteriosclerotic dementia with delirium	
Read V3 (CTV3)	E0042	Arteriosclerotic dementia with paranoia	
Read V3 (CTV3)	E0043	Arteriosclerotic dementia with depression	
Read V3 (CTV3)	E004z	Arteriosclerotic dementia NOS	
Read V3 (CTV3)	E00y.	(Oth senile/presen org psychoses) or (presbyophren psychos)	
Read V3 (CTV3)	E00z.	Senile or presenile psychoses NOS	
Read V3 (CTV3)	E011.	Korsakov psychosis	
Read V3 (CTV3)	E0110	Korsakov psychosis	
Read V3 (CTV3)	E0111	Korsakov's alcoholic psychosis with peripheral neuritis	
Read V3 (CTV3)	E0112	Wernicke-Korsakov syndrome	

Read V3 (CTV3)	E011z	Alcohol amnestic syndrome NOS	
Read V3 (CTV3)	E012.	Alcoholic dementia: [other] or [NOS]	
Read V3 (CTV3)	E040.	Korsakoff's syndrome - non-alcoholic	
Read V3 (CTV3)	E041.	Dementia in conditions EC	
Read V3 (CTV3)	Eu00.	[X]Dementia in Alzheimer's disease	x
Read V3 (CTV3)	Eu001	Dementia in Alzheimer's disease with late onset	x
Read V3 (CTV3)	Eu002	[X]Dementia in Alzheimer's dis, atypical or mixed type	x
Read V3 (CTV3)	Eu00z	[X]Dementia in Alzheimer's disease, unspecified	x
Read V3 (CTV3)	Eu01.	Vascular dementia	
Read V3 (CTV3)	Eu011	[X]Dementia: [multi-infarct] or [predominantly cortical]	
Read V3 (CTV3)	Eu01y	[X]Other vascular dementia	
Read V3 (CTV3)	Eu01z	[X]Vascular dementia, unspecified	
Read V3 (CTV3)	Eu02.	[X]Dementia in other diseases classified elsewhere	
Read V3 (CTV3)	Eu020	[X]Dementia in Pick's disease	
Read V3 (CTV3)	Eu021	[X]Dementia in Creutzfeldt-Jakob disease	
Read V3 (CTV3)	Eu022	[X]Dementia in Huntington's disease	
Read V3 (CTV3)	Eu023	[X]Dementia in Parkinson's disease	
Read V3 (CTV3)	Eu02y	[X]Dementia in other specified diseases classif elsewhere	
Read V3 (CTV3)	Eu02z	[X] Dementia: [unspecified] or [named variants (& NOS)]	
Read V3 (CTV3)	Eu041	[X]Delirium superimposed on dementia	
Read V3 (CTV3)	F110.	Alzheimer's disease	x
Read V3 (CTV3)	F1100	Dementia in Alzheimer's disease with early onset	x
Read V3 (CTV3)	F1101	Dementia in Alzheimer's disease with late onset	x
Read V3 (CTV3)	F111.	Pick's disease	
Read V3 (CTV3)	F11x7	Cerebral degeneration due to Creutzfeldt-Jakob disease	
Read V3 (CTV3)	F11x8	Cerebral degeneration due to multifocal leucoencephalopathy	
Read V3 (CTV3)	F21y2	Binswanger's disease	
Read V3 (CTV3)	Fyu30	[X]Other Alzheimer's disease	x
Read V3 (CTV3)	Ub1T6	Language disorder of dementia	
Read V3 (CTV3)	X002m	Amyotrophic lateral sclerosis with dementia	
Read V3 (CTV3)	X002w	Dementia	

Read V3 (CTV3)	X002x	Dementia in Alzheimer's disease with early onset	x
Read V3 (CTV3)	X002y	Familial Alzheimer's disease of early onset	x
Read V3 (CTV3)	X002z	Non-familial Alzheimer's disease of early onset	x
Read V3 (CTV3)	X0030	Dementia in Alzheimer's disease with late onset	x
Read V3 (CTV3)	X0031	Familial Alzheimer's disease of late onset	x
Read V3 (CTV3)	X0032	Non-familial Alzheimer's disease of late onset	x
Read V3 (CTV3)	X0033	Focal Alzheimer's disease	x
Read V3 (CTV3)	X0034	Frontotemporal dementia	
Read V3 (CTV3)	X0035	Pick's disease with Pick bodies	
Read V3 (CTV3)	X0036	Pick's disease with Pick cells and no Pick bodies	
Read V3 (CTV3)	X0037	Frontotemporal degeneration	
Read V3 (CTV3)	X0039	Frontal lobe degeneration with motor neurone disease	
Read V3 (CTV3)	X003A	Lewy body disease	
Read V3 (CTV3)	X003G	Progressive aphasia in Alzheimer's disease	x
Read V3 (CTV3)	X003H	Argyrophilic brain disease	
Read V3 (CTV3)	X003I	Post-traumatic dementia	
Read V3 (CTV3)	X003J	Punch drunk syndrome	
Read V3 (CTV3)	X003K	Spongiform encephalopathy	
Read V3 (CTV3)	X003L	Prion protein disease	
Read V3 (CTV3)	X003M	Gerstmann-Straussler-Scheinker syndrome	
Read V3 (CTV3)	X003P	Acquired immune deficiency syndrome dementia complex	
Read V3 (CTV3)	X003R	Vascular dementia of acute onset	
Read V3 (CTV3)	X003T	Subcortical vascular dementia	
Read V3 (CTV3)	X003V	Mixed cortical and subcortical vascular dementia	
Read V3 (CTV3)	X003W	Semantic dementia	
Read V3 (CTV3)	X003X	Patchy dementia	
Read V3 (CTV3)	X003Y	Epileptic dementia	
Read V3 (CTV3)	X003l	Parkinson's disease - dementia complex on Guam	
Read V3 (CTV3)	X00R0	Presbyophrenic psychosis	
Read V3 (CTV3)	X00R2	Senile dementia	
Read V3 (CTV3)	X00Rk	Alcoholic dementia NOS	

Read V3 (CTV3)	X73mf	Creutzfeldt-Jakob disease agent	
Read V3 (CTV3)	X73mj	Bovine spongiform encephalopathy agent	
Read V3 (CTV3)	XE1Xr	Senile and presenile organic psychotic conditions	
Read V3 (CTV3)	XE1Xs	Vascular dementia	
Read V3 (CTV3)	XE1Xt	Other senile and presenile organic psychoses	
Read V3 (CTV3)	XE1Xu	Other alcoholic dementia	
Read V3 (CTV3)	XE1Z6	[X]Unspecified dementia	
Read V3 (CTV3)	XE1aG	Dementia (& [presenile] or [senile])	
Read V3 (CTV3)	Xa0lH	Multi-infarct dementia	
Read V3 (CTV3)	Xa0sC	Frontal lobe degeneration	
Read V3 (CTV3)	Xa0sE	Dementia of frontal lobe type	
Read V3 (CTV3)	Xa1GB	Cerebral degeneration presenting primarily with dementia	
Read V3 (CTV3)	Xa25J	Alcoholic dementia	
Read V3 (CTV3)	Xa3ez	Other senile/presenile dementia	
Read V3 (CTV3)	XaA1S	New variant of Creutzfeldt-Jakob disease	
Read V3 (CTV3)	XaE74	Senile dementia of the Lewy body type	
Read V3 (CTV3)	XaIKB	Alzheimer's disease with early onset	x
Read V3 (CTV3)	XaIKC	Alzheimer's disease with late onset	x
Read V3 (CTV3)	XaKyY	[X]Lewy body dementia	
Read V3 (CTV3)	XaLFf	Exception reporting: dementia quality indicators	
Read V3 (CTV3)	XaLFo	Excepted from dementia quality indicators: Patient unsuitabl	
Read V3 (CTV3)	XaLFp	Excepted from dementia quality indicators: Informed dissent	
Read V3 (CTV3)	XaMAo	Prion protein markers for Creutzfeldt-Jakob disease	
Read V3 (CTV3)	XaMFy	Dementia monitoring administration	
Read V3 (CTV3)	XaMG0	Dementia monitoring first letter	
Read V3 (CTV3)	XaMGF	Dementia annual review	
Read V3 (CTV3)	XaMGG	Dementia monitoring second letter	
Read V3 (CTV3)	XaMGI	Dementia monitoring third letter	
Read V3 (CTV3)	XaMGJ	Dementia monitoring verbal invite	
Read V3 (CTV3)	XaMGK	Dementia monitoring telephone invite	
Read V3 (CTV3)	XaMJC	Dementia monitoring	

Read V3 (CTV3)	XaZWz	Participates in Butterfly Scheme for dementia	
Read V3 (CTV3)	XaZX0	Butterfly Scheme for dementia declined	
Read V3 (CTV3)	XabVp	Sporadic Creutzfeldt-Jakob disease	
Read V3 (CTV3)	XabtQ	Dementia medication review	
Read V3 (CTV3)	Xaghb	Predominantly cortical dementia	
Read V3 (CTV3)	Y000c	Dementia review done	
Read V3 (CTV3)	Y1f1d	Dementia monitoring invitation	
Read V3 (CTV3)	Y1f22	Dementia monitoring invitation	
Read V3 (CTV3)	Y1f98	Quality and Outcomes Framework dementia quality indicator-related care invitation (procedure)	
Read V3 (CTV3)	Y23fb	Mixed dementia	
Read V3 (CTV3)	Y6230	Creutzfeldt - Jakob disease	
Read V3 (CTV3)	Y8180	Other senile/presenile dement.	
Read V3 (CTV3)	Y9086	Senile dementia - simple type	
Read V3 (CTV3)	Y9087	Senile dementia-acute confused	
Read V2	Fyu3000	[X]Other Alzheimer's disease	x
Read V2	Eu01.11	[X]Arteriosclerotic dementia	
Read V2	E004300	Arteriosclerotic dementia with depression	
Read V2	Eu04100	[X]Delirium superimposed on dementia	
Read V2	E002z00	Senile dementia with depressive or paranoid fe...	
Read V2	Eu00013	[X]Alzheimer's disease type 2	x
Read V2	Eu01y00	[X]Other vascular dementia	
Read V2	Eu02y00	[X]Dementia in other specified diseases classi...	
Read V2	Eu02z11	[X] Presenile dementia NOS	
Read V2	Eu00200	[X]Dementia in Alzheimer's dis atypical or mi...	x
Read V2	E00..11	Senile dementia	
Read V2	Eu02z13	[X] Primary degenerative dementia NOS	
Read V2	E000.00	Uncomplicated senile dementia	
Read V2	E002100	Senile dementia with depression	
Read V2	E001000	Uncomplicated presenile dementia	
Read V2	Eu01000	[X]Vascular dementia of acute onset	

Read V2	Eu00012	[X]Primary degen dementia Alzheimer's type p...	x
Read V2	Eu00100	[X]Dementia in Alzheimer's disease with late o...	x
Read V2	E004.00	Arteriosclerotic dementia	
Read V2	E004000	Uncomplicated arteriosclerotic dementia	
Read V2	E001100	Presenile dementia with delirium	
Read V2	E004z00	Arteriosclerotic dementia NOS	
Read V2	Eu01300	[X]Mixed cortical and subcortical vascular dem...	
Read V2	E004200	Arteriosclerotic dementia with paranoia	
Read V2	E004.11	Multi infarct dementia	
Read V2	F112.00	Senile degeneration of brain	
Read V2	F110.00	Alzheimer's disease	x
Read V2	E00..12	Senile/presenile dementia	
Read V2	E002.00	Senile dementia with depressive or paranoid fe...	
Read V2	F110000	Alzheimer's disease with early onset	x
Read V2	E001300	Presenile dementia with depression	
Read V2	E003.00	Senile dementia with delirium	
Read V2	E001z00	Presenile dementia NOS	
Read V2	E001200	Presenile dementia with paranoia	
Read V2	Eu01200	[X]Subcortical vascular dementia	
Read V2	Eu00011	[X]Presenile dementiaAlzheimer's type	x
Read V2	Eu02z14	[X] Senile dementia NOS	
Read V2	E004100	Arteriosclerotic dementia with delirium	
Read V2	Eu02z16	[X] Senile dementia depressed or paranoid type	
Read V2	Eu00112	[X]Senile dementiaAlzheimer's type	x
Read V2	Eu00.00	[X]Dementia in Alzheimer's disease	x
Read V2	Eu00000	[X]Dementia in Alzheimer's disease with early ...	x
Read V2	F110100	Alzheimer's disease with late onset	x
Read V2	Eu02z00	[X] Unspecified dementia	
Read V2	Eu01z00	[X]Vascular dementia unspecified	
Read V2	E001.00	Presenile dementia	
Read V2	Eu01100	[X]Multi-infarct dementia	

Read V2	E002000	Senile dementia with paranoia	
Read V2	Eu01111	[X]Predominantly cortical dementia	
Read V2	Eu00113	[X]Primary degen dementia of Alzheimer's type...	x
Read V2	Eu00z11	[X]Alzheimer's dementia unspec	x
Read V2	1461	H/O: dementia	
Read V2	Eu00z00	[X]Dementia in Alzheimer's disease unspecified	x
Read V2	Eu01.00	[X]Vascular dementia	
Read V2	Eu00111	[X]Alzheimer's disease type 1	x

Read V3 codes were based on the following code list (ID 48c76cf8) from OpenCodelists:

https://www.opencodelists.org/codelist/opensafely/dementia-complete/48c76cf8/#full-list

Read V2 codes were based on code list PH859 / 1797 from the CALIBER phenotype library:

https://phenotypes.healthdatagateway.org/phenotypes/PH859/version/1797/detail/