medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Short Report: 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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

INTRODUCTION

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 the ε3 and ε4 *APOE* alleles.

METHODS

We conducted a cohort study using genetic and electronic health record data from 171,128 participants of UK Biobank, aged ≥60 years at baseline. AD and all-cause dementia were ascertained at baseline and during follow-up for up to 16.8 years.

RESULTS

74.0% (95% CI: 37.8, 89.1) of AD and 38.1% (95% CI: 9.8, 57.6%) of all-cause dementia burden were attributable to a combination of ε3 and ε4 carriage. Approximately 30% of AD was attributable to ε3 carriage specifically.

DISCUSSION

Differences in apolipoprotein E are responsible for most AD and a large fraction of all dementia. Research into this pathway should be prioritised to facilitate dementia prevention.

Keywords

Alzheimer's disease, apolipoprotein E, APOE, attributable risk, population attributable fraction, UK Biobank

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abbreviations

ApoE – apolipoprotein E

CAD – Coronary artery disease

PAF – Population attributable fraction

RR – Risk ratio

UKB – UK Biobank

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1. Introduction

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, despite ε3 substantially increasing AD risk relative to ε2 carriage.[4, 5] Previous PAF estimates for ε4 alone have also been too low for the following reasons: i) biases in study design – including ascertainment of AD/dementia solely through clinical follow-ups (with cases being prone to attrition), the elimination of prevalent cases from samples before calculation, and/or limited follow-up time in incidence studies;[2] ii) combining ε3 and ε2 carriers in a reference group, which means risk of AD due to ε4 carriage is compared to risk with carriage of both moderate and low-risk alleles; iii) erroneous calculations -- for instance, when allele frequency (an allele's proportion among *all alleles* in a sample) has been used to define the prevalence of exposure to risk genotypes in PAF estimates, instead of genotype frequency (the proportion of individuals in a sample with a given genotype).[4] Establishing the full extent of the AD burden that arises due to common differences in apoE is important because it indicates the proportion of cases that could be prevented by intervening on this single molecular pathway. Using data from the UK Biobank cohort study (UKB), we provide more comprehensive estimates of

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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 most individuals.

2. Methods

2.1 Sample

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), the random removal of one individual from related pairs, and individuals who may have had ε3r alleles of *APOE* – leaving an analytical sample of 171,128.

2.2 APOE coding

APOE ε2/ε3/ε4 alleles were coded from genotyped or hard-called imputed microarray data for single nucleotide polymorphisms rs7412 and rs429358.[6] In primary analyses, we coded *APOE* genotypes as one exposure representing ε3 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).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2.3 Outcome ascertainment

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). AD was ascertained only via record linkage and not by baseline self-reporting due to the wording of the survey item at the baseline assessment enquiring about a 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 supplementary table 1; linkage is being sought for the remainder of the cohort). In 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.

2.4 Statistical analysis

Risks of AD and all-cause dementia by *APOE* genotype were estimated using generalized linear models with a binomial distribution and log link (log-binomial modelling), which yield risk ratios (RR). Log-binomial models were adopted for analyses rather than survival analysis because of the sample's mix of prevalent and incident cases, and RRs are used in the formula for PAFs. Models were adjusted 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 UKB,[6] and genotyping array. The prevalence of the

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

exposure (genotype frequency) in the full sample and computed RRs for AD and all-cause dementia were used to calculate PAFs[7]:

$$AF\ (population) = \frac{Prevalence_{exposure}\ (RR-1)}{1+Prevalence_{exposure}\ (RR-1)} \quad (1)$$

95% confidence intervals for PAFs were derived using the lower and upper confidence intervals for the RRs. PAFs and their confidence intervals were converted from fractions to percentages.

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.[8] 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 log-binomial 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 2.42:1 for ε4:ε3, meaning that ε4 was estimated to be responsible for 70.1% of the increased risk experienced by ε3/ε4 carriers – and hence 70.1% 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.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

To provide context for the magnitude of the PAF of AD and dementia due to ε3 and ε4, we also calculated 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 ($2 \times p \times (1-p)$). The GWAS estimated odds by assuming additive effects of variants, so 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 the prevalence of the exposure in equation (1). For homozygotes, odds ratios were recalculated as exp(2 × log-odds) reported by the GWAS to reflect them having double the risk in an additive model; for heterozygotes, odds ratios were based on the reported log-odds. For a wider comparison of PAFs for AD to PAFs for loci related to another outcome, we calculated the equivalent statistics for coronary artery disease (CAD) using summary statistics from one of the largest CAD GWAS to date.[9] We used the same approach for AD risk loci to calculate PAFs for the top 50 loci identified from this GWAS of CAD, and then ranked these to use the highest 10 PAFs in Figure 1. Given that these calculations used odds rather than RRs, the PAFs for AD and CAD risk loci besides *APOE* may have been slightly inflated. However, the deviation between odds and risk is minor for modest association magnitudes, and no odds ratio used in these calculations exceeded 1.21.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2.5 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.

3. Results

Sample characteristics are displayed in Table 1. 99.4% had ε3 or ε4 carriage. The risk ratio for AD due to ε3 or ε4 carriage was 3.86 (95% CI: 1.61, 9.26; Table 2, Section 1). The equivalent for all-cause dementia was 1.62 (1.11, 2.37). Together, the two risk alleles accounted for 74.0% (37.8, 89.1%) and 38.1% (9.8, 57.6%) for AD and all-cause dementia, respectively. In sensitivity analyses using ascertainment of AD and dementia without diagnoses from primary care records (supplementary table 2), the PAFs were slightly lower for AD (73.1%; 35.7, 88.8%) and slightly higher for all-cause dementia (40.8%; 12.4, 60.1%).

In extended PAF calculations estimating the individual contributions of ε3 and ε4 to the overall burden of AD, 43.7% of AD burden was attributable to ε4 carriage (Table 2, Section 2). This consisted of contributions from ε2/ε4 (1.9%), ε4/ε4 (12.9%) and a share of the PAF for ε3/ε4 carriers (28.9%). The remainder of the overall PAF for AD (30.2%) was attributable to ε3 carriage specifically, with contributions from ε2/ε3 (2.1%), ε3/ε3 (16.2%) and the remaining share of the PAF due to ε3/ε4 carriage (11.9%). Note that the sum of these fractions for ε3 and ε4 separately (73.9%) differs slightly to the point estimate in our main analysis (74.0%) due to differences in precision when making five individual comparisons, rather than the use of one binary exposure.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Comparing the total PAF for ε3 and ε4 carriage to other genetic risk loci for AD, and to risk loci for CAD, no other PAF for either disease exceeded 22% (Figure 1). The preponderance of AD burden attributable to genetic variation at a single locus is exceptional among common, complex chronic diseases.

4. Discussion

Our findings indicate that if interventions could eliminate 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. Such a magnitude of attributable risk has been suggested previously (ε3 and ε4 carriage perhaps accounting for 95% of AD[10]) but not demonstrated directly before. Very large analytical samples are required to use rare ε2 homozygotes as the reference group in analyses, and there has been little recognition among dementia researchers that the ε3 allele should also be considered risk-increasing for AD. Nonetheless, we estimate that the ε3 allele alone could be responsible for almost a third or more of AD due to ε3 conferring considerable risk to most individuals. It should be noted that PAFs are distinct from heritability, with the heritability of AD estimated to range up to 79%.[11] However, heritability analyses are not informative for assessing disease burden attributable to specific causes, and PAFs are more appropriate and intuitive for this purpose.[12] We also emphasise that our results are not suggesting that AD is monogenic. AD is clearly multifactorial, but its other component causes beyond apoE are of much less consequence without the background of risk from ε3 or ε4 carriage that most people inherit. Put differently, if all individuals inherited an ε2/ε2 genotype, most AD would not occur, regardless of what other AD risk factors people experience.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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)[13] 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. Our analyses may have been affected by selection into the study or survival,[14] perhaps including effects from *APOE* genotypes contributing to cardiovascular and other morbidity and mortality before age 60 years.[15] 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.[16] However, these biases mean our calculations are likely underestimates. Attributable fractions also assume that the exposure of interest is a cause of the disease being investigated and not based on biased estimates of causation (e.g. due to confounding or reverse causation), which is often an unrealistic assumption for phenotypic traits.[7] However, 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.[17] 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.[18] Hence, causation between *APOE* variants and AD is not equivocal and 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.[19]

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Therefore, considering that most AD could be prevented (or at least delayed) by modulating the risk conferred by apoE, understanding the protein's detrimental effects should be given proportionate research attention and funding. These should include efforts to understand the distinct functional properties of the ε3 isoform responsible for AD risk – relative to properties of the ε2 isoform and other protective variants [20, 21] – and not only further research to elucidate and mitigate ε4's effects.

There is considerable scope to target apoE with interventions. With gene editing, therapy and silencing, genetic risk is now directly modifiable. Moreover, many strategies exist to target apoE at the protein level or its molecular intermediaries, including immunotherapy and small molecule structural correctors.[22, 23] However, only one therapy targeting apoE (LX1001)[24] is currently being trialled for AD in humans – less than 1% of potential therapies in registered trials.[25] 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 does not preclude investigations into other genetic or environmental factors that could be mediating or modifying the effects of apoE on AD or research into factors that may be distinct contributing causes of these outcomes (both scenarios include research addressing cerebral amyloidosis and tauopathy). Nonetheless, establishing how, when, and in which cell types apoE influences AD risk – and how its deleterious effects can be mitigated – is clearly paramount to AD prevention and treatment.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 71702. This work uses data provided by patients and collected by the NHS as part of their care and support. Copyright © 2023, NHS England. Re-used with the permission of the NHS England and UK Biobank. All rights reserved.

Funding

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).

Data Availability

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 will be made available at the following site upon publication of this article: https://github.com/dylwil/ad_apoe_paf

Competing interests

We have no competing interests to disclose.

Author contributions

DMW conceived the study, undertook the analyses and drafted the manuscript. All authors contributed to the interpretation of data and the content of the manuscript,

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

and approved its final version. The funders had no direct role in the production of this research.

References

[1] 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68.

[2] Ashford JW. APOE genotype effects on Alzheimer's disease onset and epidemiology. Journal of Molecular Neuroscience. 2004;23:157-65.

[3] Saddiki H, Fayosse A, Cognat E, Sabia S, Engelborghs S, Wallon D, et al. Age and the association between apolipoprotein E genotype and Alzheimer disease: a cerebrospinal fluid biomarker–based case–control study. PLoS Medicine. 2020;17:e1003289.

[4] Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013;45:1452-8.

[5] Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. Jama. 1997;278:1349-56.

[6] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203-9.

[7] Steenland K, Armstrong B. An overview of methods for calculating the burden of disease due to specific risk factors. Epidemiology. 2006:512-9.

[8] Hanley J. A heuristic approach to the formulas for population attributable fraction. Journal of Epidemiology & Community Health. 2001;55:508-14.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[9] A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nature genetics. 2015;47:1121-30.

[10] Raber J, Huang Y, Ashford JW. ApoE genotype accounts for the vast majority of AD risk and AD pathology. Neurobiology of aging. 2004;25:641-50.

[11] Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Archives of general psychiatry. 2006;63:168-74.

[12] 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. 2019;34:987-92.

[13] Wilkinson T, Ly A, Schnier C, Rannikmäe K, Bush K, Brayne C, et al. Identifying dementia cases with routinely collected health data: a systematic review. Alzheimer's & Dementia. 2018;14:1038-51.

[14] Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. American Journal of Epidemiology. 2017;186:1026-34.

[15] Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. 2017;49:1385-91.

[16] Reiman EM, Arboleda-Velasquez JF, Quiroz YT, Huentelman MJ, Beach TG, Caselli RJ, et al. Exceptionally low likelihood of Alzheimer’s dementia in APOE2 homozygotes from a 5,000-person neuropathological study. Nature Communications. 2020;11:667.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[17] Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. International journal of epidemiology. 2004;33:30-42.

[18] Belloy ME, Napolioni V, Greicius MD. A quarter century of APOE and Alzheimer's disease: progress to date and the path forward. Neuron. 2019;101:820-38.

[19] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet. 2020;396:413-46.

[20] Le Guen Y, Belloy ME, Grenier-Boley B, De Rojas I, Castillo-Morales A, Jansen I, et al. Association of rare APOE missense variants V236E and R251G with risk of Alzheimer disease. JAMA neurology. 2022;79:652-63.

[21] Arboleda-Velasquez JF, Lopera F, O'Hare M, Delgado-Tirado S, Marino C, Chmielewska N, et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. Nature medicine. 2019;25:1680-3.

[22] Marino C, Perez-Corredor P, O'Hare M, Heuer A, Chmielewska N, Gordon H, et al. APOE Christchurch-mimetic therapeutic antibody reduces APOE-mediated toxicity and tau phosphorylation. Alzheimer's & Dementia. 2023.

[23] Wang C, Najm R, Xu Q, Jeong D-e, Walker D, Balestra ME, et al. Gain of toxic apolipoprotein E4 effects in human iPSC-derived neurons is ameliorated by a small-molecule structure corrector. Nature medicine. 2018;24:647-57.

[24] Rosenberg JB, Kaplitt MG, De BP, Chen A, Flagiello T, Salami C, et al. AAVrh. 10-mediated APOE2 central nervous system gene therapy for APOE4-associated Alzheimer's disease. Human Gene Therapy Clinical Development. 2018;29:24-47.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[25] Cummings J, Zhou Y, Lee G, Zhong K, Fonseca J, Cheng F. Alzheimer's disease drug development pipeline: 2023. Alzheimer's & Dementia: Translational Research & Clinical Interventions. 2023;9:e12385.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1. Descriptive characteristics of 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/ε2		1,053 (0.6)
ε2/ε3		21,009 (12.3)
ε2/ε4		4,182 (2.4)
ε3/ε3		100,892 (59.0)
ε3/ε4		39,954 (23.4)
ε4/ε4		4,038 (2.4)
ε2 carriage		26,244 (15.3)
ε3 carriage		161,855 (94.6)
ε4 carriage		48,174 (28.2)
Alzheimer's disease		3,126 (1.8)
All-cause dementia		6,859 (4.0)

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[a] 'Other' refers to anyone who did not select another ethnic group

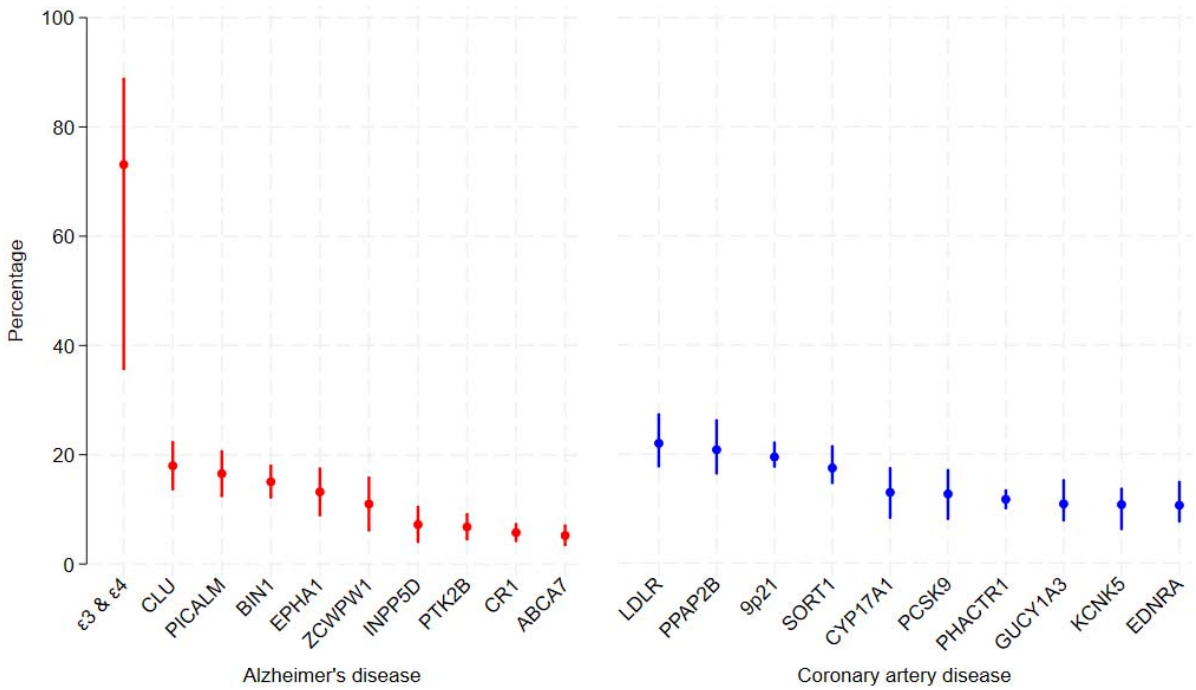
[b] NB: carriage groups overlap and hence percentages sum to over 100%

Table 2. Risk ratios (RR) and population attributable fractions (PAF) for AD or all-cause dementia by *APOE* genotypes

Exposure	N	AD				All-cause dementia			
		RR	95% CI	P	PAF, % (95%CI)	RR	95% CI	P	PAF, % (95%CI)
1. With ε3 and ε4 grouped									
ε2/ε2	1053	Ref.				Ref.			
ε3 or ε4 carriage	170075	3.86	(1.61, 9.26)	0.002	74.0 (37.8, 89.1)	1.62	(1.11, 2.37)	0.01	38.1 (9.8, 57.6)
2. By individual genotype									
ε2/ε2	1053	Ref.				Ref.			
ε2/ε3	21009	1.65	(0.68, 4.02)	0.27	2.1 (0.0, 4.0)	0.91	(0.62, 1.34)	0.64	0.0 (0.0, 1.8)
ε2/ε4	4182	4.01	(1.63, 9.86)	0.003	1.9 (1.0, 2.4)	1.88	(1.26, 2.81)	0.002	1.3 (0.6, 1.9)
ε3/ε3	100892	2.05	(0.85, 4.92)	0.11	16.2 (0.0, 25.1)	1.07	(0.73, 1.56)	0.73	2.6 (0.0, 14.2)
ε3/ε4	39954	7.68	(3.20, 18.43)	<0.001	40.8 (32.2, 44.3)	2.78	(1.90, 4.06)	<0.001	25.8 (19.1, 30.4)
ε4/ε4	4038	21.97	(9.13, 52.87)	<0.001	12.9 (12.0, 13.3)	6.96	(4.74, 10.21)	<0.001	8.7 (8.0, 9.2)

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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 determining circulating low-density lipoprotein cholesterol concentrations and downstream risk of coronary artery disease, for example.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.16.23298475; this version posted April 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .