**Supplementary Material**

**Supplementary Materials and Methods**

*Sample*

MoBa, ND, CPLS and NMB were merely observational studies, whereas the NCP study included one or more 10 week periods of memory training or rest, with scans in between these. All persons in the NCP project were offered memory training, irrespective of genetics, and hence, this should not constitute a bias. The sample is partly overlapping with a lifespan sample used in previous reports 1, 2. No relations between polygenetic risk and hippocampal volumes have been reported for any part of this sample previously, and *APOE* results in relation to hippocampal volume has only been reported for the NBM subsample (n = 86 at Tp1).3

The participants in the sub-studies run by LCBC4-7, constituting the majority (n = 1095), were cognitively healthy community dwellers, but complete absence of health problems was not required for inclusion. Participants with common health conditions, such as moderately elevated blood pressure and being on hypertensive treatment, were not excluded. They were recruited in part by newspaper and online ads, and in part through the population registry cohort study MoBa5. The participants in the NBM study3 (n = 86 at baseline) were recruited among patients scheduled for elective gynecological (genital prolapse), urological (benign prostate hyperplasia, prostate cancer, or bladder tumor/cancer) or orthopedic (knee or hip replacement) surgery in spinal anesthesia, turning 65 years or older in the year of inclusion. Additional criteria for being included in the present analyses were 1) having valid genetic data, 2) being of European ancestry as determined by genetic analyses (see below), 3) having at least one valid anatomical MRI scan with successful automatic hippocampal segmentation (see below).

Education was recorded as number of years of education to the highest attained degree for adults (age >= 18 years), and for participants below 18 years of age, the average of paternal and maternal years of education to the highest attained degree was entered, or if unavailable, for one parent (either available). By this measure, education was obtained in a comparable manner for most participants (n = 1000, mean = 16.3 years, SD = 2.8 years, range 7-24 years). All participants were compensated a modest sum for their participation, depending on amount of examinations (for the structural scan session around NOK 500, or USD $60).

**Supplementary Table 1** **Distribution of number of participants from different sub-studies and scanners across time points** (Tp).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Tp 1** | **Tp 2** | **Tp 3** | **Tp 4** | **Tp5** | **Tp6** | **Tp7** |
| **Sub-study** |  |  |  |  |  |  |  |
| MoBa | 223 | 195 | 105 | - | - | - |  |
| NBM | 86 | 81 | 62 | 27 | - | - |  |
| ND | 224 | 130 | 63 | - | - | - |  |
| CPLS | 462 | 102 | 67 | 22 | - | - |  |
| NCP | 186 | 169 | 145 | 131 | 117 | 87 | 6 |
| **Scanner** |  |  |  |  |  |  |  |
| Avanto 1 | 416 | 352 | 123 | - | - | - | - |
| Avanto 2 | 86 | 81 | 62 | 27 | - | - | - |
| Prisma | 100 | 37 | 105 | - | - | - | - |
| Skyra | 579 | 207 | 152 | 153 | 117 | 87 | 6 |

*DNA handling, genotyping, and data processing*

*DNA extraction, QC and genotyping*. Prior to genotyping, DNA was extracted either in Oslo from saliva or buccal swabs (n=592), or at the LIGA laboratory at University of Lübeck for saliva samples (n=416, incl. 2 technical duplicates) and buccal swabs (n=586, incl. 1 technical duplicate). All DNA samples were then quantified, quality controlled, normalized, and aliquoted (to approx. 35ul at ~50ng/ul) in Lübeck leading to a total of 1,594 DNA samples (incl. 3 technical duplicates) that were subjected to genotyping using the Global Screening Array (GSA; Illumina, Inc.) with shared custom content. Genotyping was performed in two batches (batch 1: n = 1,402, batch 2: n = 192) at the Institute of Clinical and Molecular Biology at UKSH Campus Kiel on an iScan instrument following the manufacturer’s instructions. Three samples from batch 1 (2 buccal swabs, 1 saliva) failed genotyping leaving n=1,591 for further processing.

*Post-genotyping data processing, QC and imputation*. All data processing steps were performed in the LIGA laboratory in Lübeck. Genotype calling was performed in GenomeStudio v2.0.4 using manifest “GSAsharedCUSTOM\_20018389\_A6” (Illumina, Inc.) providing annotations for a total of 696,375 variants. GenTrain v3.0 (Illumina, Inc.) was used for automatic clustering and genotype calling. Quality assessments at this stage revealed one sample failing QC from batch 2 (using call rate < .95 and p50 GC < .7 as threshold) so that all 1,590 samples were exported using the PLINK Input Report 2.1.4 module of GenomeStudio. Subsequent data processing used an automated workflow developed in LIGA executed in the high-performance computing environment (“OmicsCluster”) available at University of Lübeck. This entailed exclusion of 110,579 variants with GenTrain values <.708 in a reference dataset of ~20,000 DNA samples genotyped in a separate project, conversion of variant alleles to forward (plus) stand using PLINK (v1.90b4;9; command: '--flip') and checking for inconsistencies between reported and genetic sex ('--check-sex'). At this stage, 13 samples (batch 1: n=11, batch 2: n=3) were excluded because genetic sex could not be determined unambiguously from the genotype data, and non-matching sex information was recorded for re-inspection of clinical data for 29 samples (batch 1: n = 25, batch 2: n=4). Subsequent sample- and variant-level QC entailed filtering with PLINK commands '--mind 0.05 --geno 0.02 --hwe 0.000005 --maf 0.01', resulting in 446,837 high-quality variants with MAF ≥1% across 1,562 (batch 1: n=1,381, batch 2: n=181) samples. These data were used to create an LD pruned dataset with '--indep-pairwise 1500 150 0.2 --maf 0.05' followed by pairwise genetic similarity analyses using '--Z-genome --min 0.06' to identify cryptic relatedness. All three technical replicates were identified correctly and showed PI\_HAT values of 1.0. In addition, samples with >3 standard deviations of pairwise matches at PI\_HAT>0.06 were excluded (batch 1: n=2 samples, batch 2: n=0). The LD pruned dataset was then used for principal component analysis (PCA; using PLINK command ‘--pca') along with the reference dataset of the 1000 Genomes Project Phase 3 (1000G10) to assign ethnic descent groups using the five 1000G super-populations by k-nearest neighbor (k-NN; k=9) classification (using R package ‘class’ in R 2.3.2;11). Subsequently, data were recoded into VCF format ('--recode vcf-iid'), whereby ambiguous SNPs were removed, and strand mismatches re-checked and corrected with BCFtools 1.9 ('+fixref -m flip -d -f GRCh37.fasta') and confirmed ('+af-dist'). SHAPEIT2 (v2.r837;12) was used for phasing. Variants not matching to the reference haplotypes were excluded ('-check -M), all remaining variants were phased with the same reference data, i.e. the genetic map of the 1000G and the HRC reference panel haplotypes Release 1.1 (EGAD00001002729). The phased genotype data were then subjected to imputation using the HRC reference with Minimac313 applying default parameters. Overall, this procedure resulted in 39,131,578 genotypes across 1,560 (batch 1: n=1,379, batch 2: n=181) (incl. all 3 technical replicates) samples.

*Polygenic score computation.* Imputed dosages were further quality controlled by: 1.) removing non-European subjects determined by the PCA analyses; 2.) SNPs having imputation R-square <0.8; and, 3.) minor allele frequencies (MAF) <0.05. The resultant dosage genotypes were converted to best-guess genotypes, i.e. 0, 1, or 2 copies of the minor allele for each SNP. In total, 5.2 million SNPs remain. AD GWAS summary statistics were downloaded from <http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php> with permission. AD-PGS of our sample was calculated using the allelic effect sizes from Lambert et al.14. Shared SNPs between our best-guess genotype dataset and the GWAS summary statistics were pruned to be near independent with PLINK using parameters --clump-p1 1.0 --clump-p2 1.0 --clump-kb 500 --clump-r2 0.1 with LD structure from 1000G. The linkage disequilibrium (LD) structure was based on the European subpopulation from the 1000 Genomes Project Phase310. To avoid the impact of the complex LD structure of the MHC region (build hg19; chr6:25652429-33,368,333), we removed SNPs in this region except the most significant one before pruning. Previous studies15-17 have shown that PGS constructed using SNPs with association p value < 0.5 from Lambert et al.14 have the largest effect on the risk of AD. Hence, we used the same threshold in the pruned set for computing the AD-PGS. To investigate consistency of results across different p-value thresholds, SNPs with p < 5e-08 in the pruned set were additionally used for constructing a PGS for our samples. Analyses were recomputed limiting markers to those showing genome-wide significant association (i.e. p < 5e-08) with AD risk in Lambert et al.14. Due to its known large effect, we computed AD-PGS with and without markers in the *APOE* region (build hg19; chr19:44,909,011-45,912,650). To test the effect of *APOE* itself we modelled the counts of *APOE* ε4 alleles directly by determining haplotypes of the two SNPs rs7412 and rs42935818, 19, coded as 0, 1, or 2 copies of the ε4 allele.

To control the possible effect of vascular factors on the relation between AD-PGS and hippocampus volume, we included PGSs for systolic and diastolic blood pressure (PGS-SBP and PGS-DBP) as additional covariates in our models. Both PGS-SBP and PGS-DBP were computed using the same protocol as that used for AD-PGS, but based on the public data from Liu et al. 20. We included SNPs having an association p value < 0.5 for computing SBP-PGS and DBP-PGS, i.e. the same as for AD-PGS. However, Liu et al. 20 only provided association statistics for SNPs in the exome regions, and fewer SNPs were included for computing SBP-PGS and DBP-PGS than for AD-PGS (8389 SNPs for DBP and 8328 for SBP). In addition, SNPs in the APOE gene region were excluded in both SBP-PGS and DBP-PGS.

We computed the genetic ancestry factors (GAFs) using principal components methods21. The pre-imputation QC’ed genotypes were used for estimating GAF for European samples (as determined by the k-NN approach on 1000G superpopulations, see above). SNPs with MAF < 0.1 were excluded first and then pruned to be nearly independent by PLINK using parameters, --indep-pairwise 100 50 0.1. Then, the remaining SNPs were included in the GAF estimation with the PLINK command, --pca. The top 20 principal components were retained to be used as covariates in the statistical analyses. For the present analysis, only participants of European ancestry were included, excluding 89 persons for whom we had genotype data.

## *MRI data acquisition*

Participants were scanned at a total of 4 Siemens scanners at 2 sites (1: Oslo University Hospital, 2: Curato (Currently Aleris), Oslo): A 1.5T Avanto equipped with a 12 channel head coil (Site 1 and 2), a 3T Skyra equipped with a 24-channel Siemens head coil (Site 1) or a 3T Prisma equipped with a 32 channel head coil (Site 1) (all Siemens Medical Systems, Erlangen, Germany). The pulse sequence used for morphometric analyses were one to two 3D sagittal T1-weighted MPRAGE sequences. Avanto site 1 and 2: 160 slices, repetition time (TR), 2400ms; echo time (TE), 3.61ms/3.79ms (Site 1/2); time to inversion, 1000ms; flip angle, 8**°**; matrix, 192x192; field of view, 240; voxel size, 1.25x1.25x1.20 mm per participant per visit. Scanning time for each MPRAGE sequence was 7min 42s. Skyra: 176 slices, TR = 2300 ms, TE = 2.98 ms, flip angle = 8°, voxel size = 1 × 1 × 1 mm, FOV= 256 × 256 mm. Prisma: 208 slices, TR = 2400 ms, TE = 2.22 ms, TI = 1000 ms, flip angle = 8°, voxel size = 0.8 x 0.8 x 0.8 mm3, FOV = 240 x 256 mm2. Other MRI volumes were recorded including sequences intended for and examined by a radiologist, to rule out and medically follow up incidental neuroradiological findings. Distribution of scans from the different scanners per timepoint is given in Table 2.

## 

## *Image analysis*

All scans were reviewed for quality and automatically corrected for spatial distortion22. Images were first automatically processed cross-sectionally for each time point with the FreeSurfer software package (version 6.0). This processing includes motion correction, removal of non-brain tissue, automated Talairach transformation, intensity correction and automatic volumetric segmentation, including hippocampal volumetric segmentation23, 24. In older subjects, FreeSurfer is shown to calculate consistent hippocampal volumes with reproducibility errors of 3.4%- 3.6%25. To extract reliable longitudinal subcortical volume estimates, the images were run through the longitudinal stream in FreeSurfer26, 27. Specifically, an unbiased within-subject template volume based on the cross-sectional images was created for each participant, and processing of all time points was then initialized using common information from this template. This increased sensitivity and robustness of the longitudinal analysis and ensured inverse consistency26. As part of this processing, a fused segmentation created for each time point by an intensity based probabilistic voting scheme was applied to further reduce the variability across time points. Participants followed-up on different MRI scanners were independently processed for each scanner. To allow assessment of differences between scanners, 24 participants were scanned on all three scanners from Oslo University Hospital on the same day. Linear regression analyses were run testing the concordance between hippocampal volumes between scanners, yielding excellent agreement (Avanto vs Prisma R2 = .93; Prisma vs Skyra R2 = .94; Prisma vs. Avanto R2=.90). Thus, including scanner as covariate in the analyses would account for offset differences between scanners (see 28).

*Statistical analyses*

Analyses were run in R29 version 3.6.0*.* General Additive Mixed Models (GAMM) using the package “mgcv”30 version 1.8-28 were used to derive age-functions with a random intercept term per participant.

**Supplementary Results**

*PGSs for blood pressure*

Adding PGSs for blood pressure (DBP and SBP) as additional covariates, the effect of AD-PGSon hippocampal volume was still significant. The estimated effect of AD-PGS on hippocampal volume, was then -35.8 mm3 (CI: -71.2, -0.45) associated with one sample SD higher AD-PGS, whereas the effects associated with one sample SD higher blood pressure PGSs on hippocampal volume did not reach statistical significance (DBP-PGS: -7.1 mm3, CI: -50.8, 36.6; SBP-PGS: -30.2 mm3 (CI: -74.9, 14.5).

Adding PGSs for blood pressure (DBP and SBP) as additional covariates, did not weaken the effect of *APOE* status on hippocampal volume, the estimated effect of carrying one or more ε4 alleles on hippocampal volume with these additional covariates, was -111.2 mm3 (CI: -186.7, -35.8), whereas the effects associated with one sample SD higher blood pressure PGSs on hippocampal volume did not reach statistical significance (DBP-PGS: -7.1 mm3, CI: -50.8, 36.6); SBP: -30.2 mm3 (CI: -74.9, 14.5).

*Analyses by hemisphere*

The estimated effect of AD-PGS on hippocampal volume, was in the left hemisphere -18.9 mm3 (CI: -37.3, -0.48) associated with one sample SD higher AD-PGS, and in the right hemisphere -17.4 mm3 (CI: -36.4, -1.62). A somewhat stronger interaction between age and AD-PGS was seen in the left hemisphere compared to the right hemisphere (see Supplementary Figure 1). At the age of 20, this interaction term estimated an additional offset effect -0.59 mm3 (CI: -1.52, 0.35) in the left hemisphere, and 0.17 mm3 (CI: -0.50, 0.84). in the right hemisphere at AD-PGS one standard deviation below the mean and 0.55 mm3 (CI: -0.39, 1.49) in the left hemisphere and -0.17 mm3 (CI: -0.84, 0.50) in the right hemisphere at one standard deviation above the mean, and at the age of 80 the corresponding estimates were (for -1SD) 0.46 mm3 (CI: -0.98, 1.90) in the left hemisphere and -0.09 mm3 (CI: -1.27, 1.10) in the right hemisphere, and (for +1SD) -0.36 mm3 (CI: -1.81, 1.09) in the left hemisphere and 0.09 mm3 (CI: -1.10, 1.28) in the right hemisphere.

The estimated effect of carrying one or more ε4 alleles on hippocampal volume, was -48.0 mm3 (CI: -87.0, -8.71) in the left hemisphere and -59.0 mm3 (CI: -99.5, -18.4) in the right hemisphere. The interaction effect is shown in Supplementary Figure 2. At age 20, the difference in hippocampal volume between carriers and non-carriers was -51.5 mm3 (CI: -117, 13.7) in the left hemisphere and -41.8 mm3 (CI: -110, 26.0) in the right hemisphere, at age 50 this difference was -23.4 mm3 (CI: -109, 62.3) in the left hemisphere and -10.0 mm3 (CI: -101, 80.7) in the right hemisphere, and at age 80 it was -99.8 mm3 (CI: -208, -8.08) in the left hemisphere and -169 mm3 (CI: -279, -58.6) in the right hemisphere.

*Analyses with AD-PGS limited to SNPs only showing genome-wide significant association*

Recomputing the analyses with an AD-PGS limited to SNPs only showing genome-wide significant association (i.e. p < 5e-08) with AD risk in Lambert et al.14, confirmed a significant negative effect of AD-PGS on hippocampal volume (-44.8 mm3/AD-PGS SD, CI: -80.7, -8.9). This result was reduced to a trend when excluding markers in the *APOE* region (-35.6 mm3/PGS\_SD, CI: -76.7, 5.5). However, significant age interactions appeared in the PGS analyses limited to genome-wide significant markers, both with and without markers in the *APOE* region (see Figure 2). Especially when excluding the *APOE* region, there seemed to be a somewhat more negative effect of higher AD-PGS on hippocampal volume in older age (above 80 years): At age 80, the interaction effect of AD-PGS was 2.9 mm3 (CI: 0.4, 5.4) for values one standard deviation below sample mean, and -2.8 mm3 (CI: -5.3, -0.3) for values one standard deviation above the sample mean. For AD-PGS including the APOE region, at age 80, an AD-PGS value 1 SD below mean was associated with an offset in hippocampal volume of 2.8 mm3 (CI: 0.4, 5.2), while AD-PGS value 1 SD below mean was associated with an offset of -2.8 mm3 (CI: -5.2, -0.4), compared to having a value of AD-PGS at the sample mean. Compared to the main effect of AD-PGS, estimated to -44.8 mm3, the interaction is hence small, even at the age of 80. For AD-PGS excluding the APOE region, the corresponding numbers were 2.9 mm3 (CI: 0.4, 5.4) for AD-PGS 1 SD below sample mean and -2.8 mm3 (CI: -5.3, -0.3) for AD-PGS 1SD above sample mean. At age 20, having AD-PGS (including APOE) 1 SD below sample mean was associated with an offset in hippocampal volume of -0.2 mm3 (CI: -1.7, 1.3) and having AD-PGS 1SD above sample mean with an offset 0.2 mm3 (-1.3, 1.7). The corresponding numbers excluding APOE were -0.7 mm3 (CI: -2.4, 1.0) and 0.7 mm3 (-1.0, 2.4).

**Supplementary References**

1. Fjell AM, Grydeland H, Krogsrud SK, et al. Development and aging of cortical thickness correspond to genetic organization patterns. Proceedings of the National Academy of Sciences of the United States of America 2015;112:15462-15467.

2. Walhovd KB, Krogsrud SK, Amlien IK, et al. Neurodevelopmental origins of lifespan changes in brain and cognition. Proceedings of the National Academy of Sciences of the United States of America 2016;113:9357-9362.

3. Idland AV, Sala-Llonch R, Borza T, et al. CSF neurofilament light levels predict hippocampal atrophy in cognitively healthy older adults. Neurobiology of aging 2017;49:138-144.

4. de Lange AG, Brathen ACS, Rohani DA, Fjell AM, Walhovd KB. The Temporal Dynamics of Brain Plasticity in Aging. Cerebral cortex 2018.

5. Krogsrud SK, Fjell AM, Tamnes CK, et al. Changes in white matter microstructure in the developing brain--A longitudinal diffusion tensor imaging study of children from 4 to 11years of age. NeuroImage 2016;124:473-486.

6. Storsve AB, Fjell AM, Tamnes CK, et al. Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: regions of accelerating and decelerating change. The Journal of neuroscience : the official journal of the Society for Neuroscience 2014;34:8488-8498.

7. Tamnes CK, Walhovd KB, Grydeland H, et al. Longitudinal working memory development is related to structural maturation of frontal and parietal cortices. Journal of cognitive neuroscience 2013;25:1611-1623.

8. Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. Nat Protoc 2014;9:2643-2662.

9. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.

10. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. Nature 2015;526:68-74.

11. Venables WN, Ripley BD. Modern Applied Statistics with S, 4th Edition ed: Springer, Boston, MA, 2002.

12. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nature methods 2012;9:179-181.

13. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-1287.

14. Lambert JC, 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;45:1452-1458.

15. Escott-Price V, Myers AJ, Huentelman M, Hardy J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Annals of neurology 2017;82:311-314.

16. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. Brain : a journal of neurology 2015;138:3673-3684.

17. Logue MW, Panizzon MS, Elman JA, et al. Use of an Alzheimer's disease polygenic risk score to identify mild cognitive impairment in adults in their 50s. Molecular psychiatry 2019;24:421-430.

18. Laws SM, Hone E, Gandy S, Martins RN. Expanding the association between the APOE gene and the risk of Alzheimer's disease: possible roles for APOE promoter polymorphisms and alterations in APOE transcription. J Neurochem 2003;84:1215-1236.

19. Radmanesh F, Devan WJ, Anderson CD, Rosand J, Falcone GJ, Alzheimer's Disease Neuroimaging I. Accuracy of imputation to infer unobserved APOE epsilon alleles in genome-wide genotyping data. Eur J Hum Genet 2014;22:1239-1242.

20. Liu C, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nat Genet 2016;48:1162-1170.

21. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190.

22. Jovicich J, Czanner S, Greve D, et al. Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. NeuroImage 2006;30:436-443.

23. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341-355.

24. Han X, Jovicich J, Salat D, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage 2006;32:180-194.

25. Jovicich J, Czanner S, Han X, et al. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. NeuroImage 2009;46:177-192.

26. Reuter M, Rosas HD, Fischl B. Highly accurate inverse consistent registration: a robust approach. NeuroImage 2010;53:1181-1196.

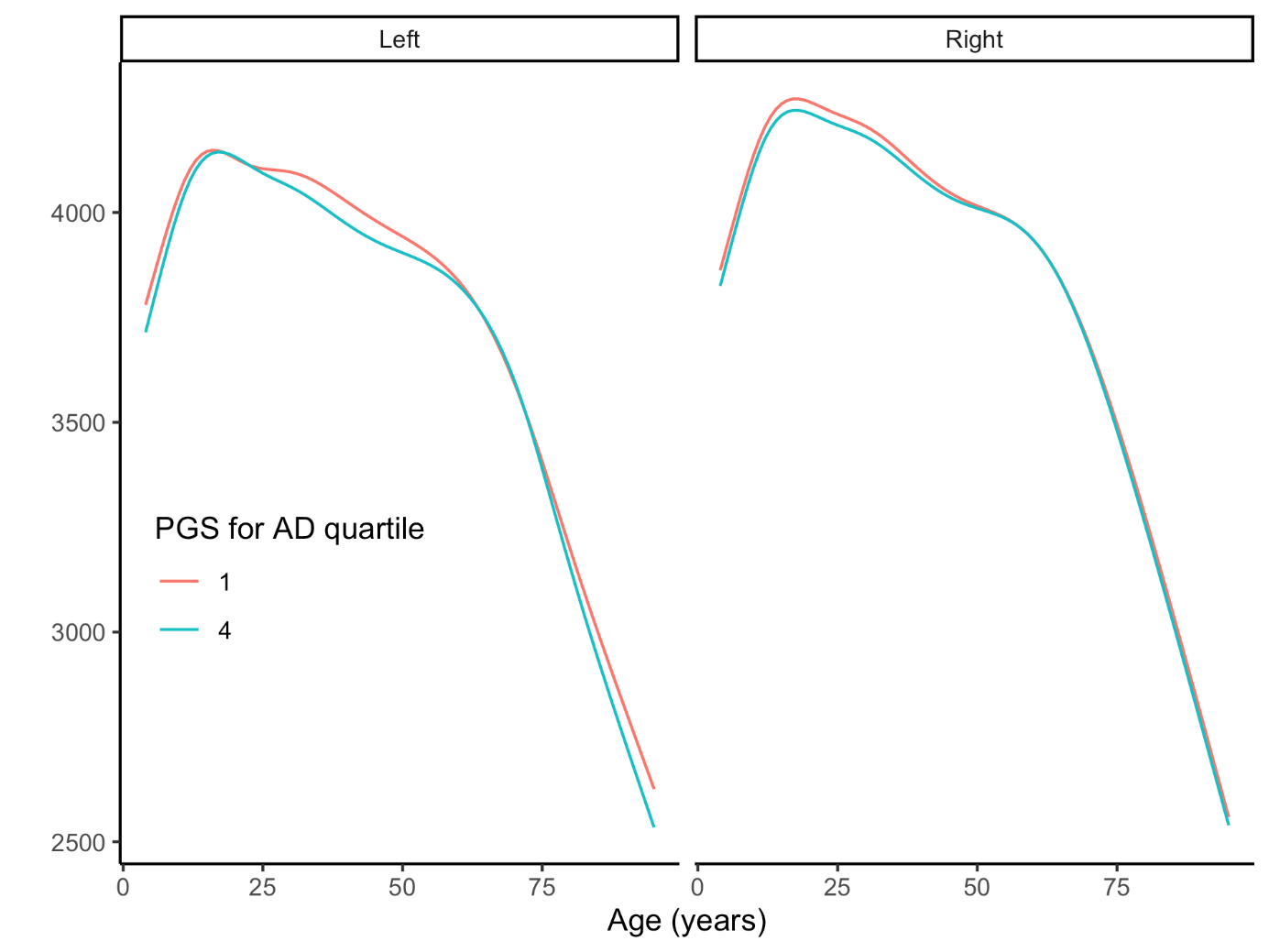
27. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. NeuroImage 2012;61:1402-1418.

28. Nyberg CK, Fjell AM, Walhovd KB. Level of body fat relates to memory decline and interacts with age in its association with hippocampal and subcortical atrophy. Neurobiology of aging 2020;91:112-124.

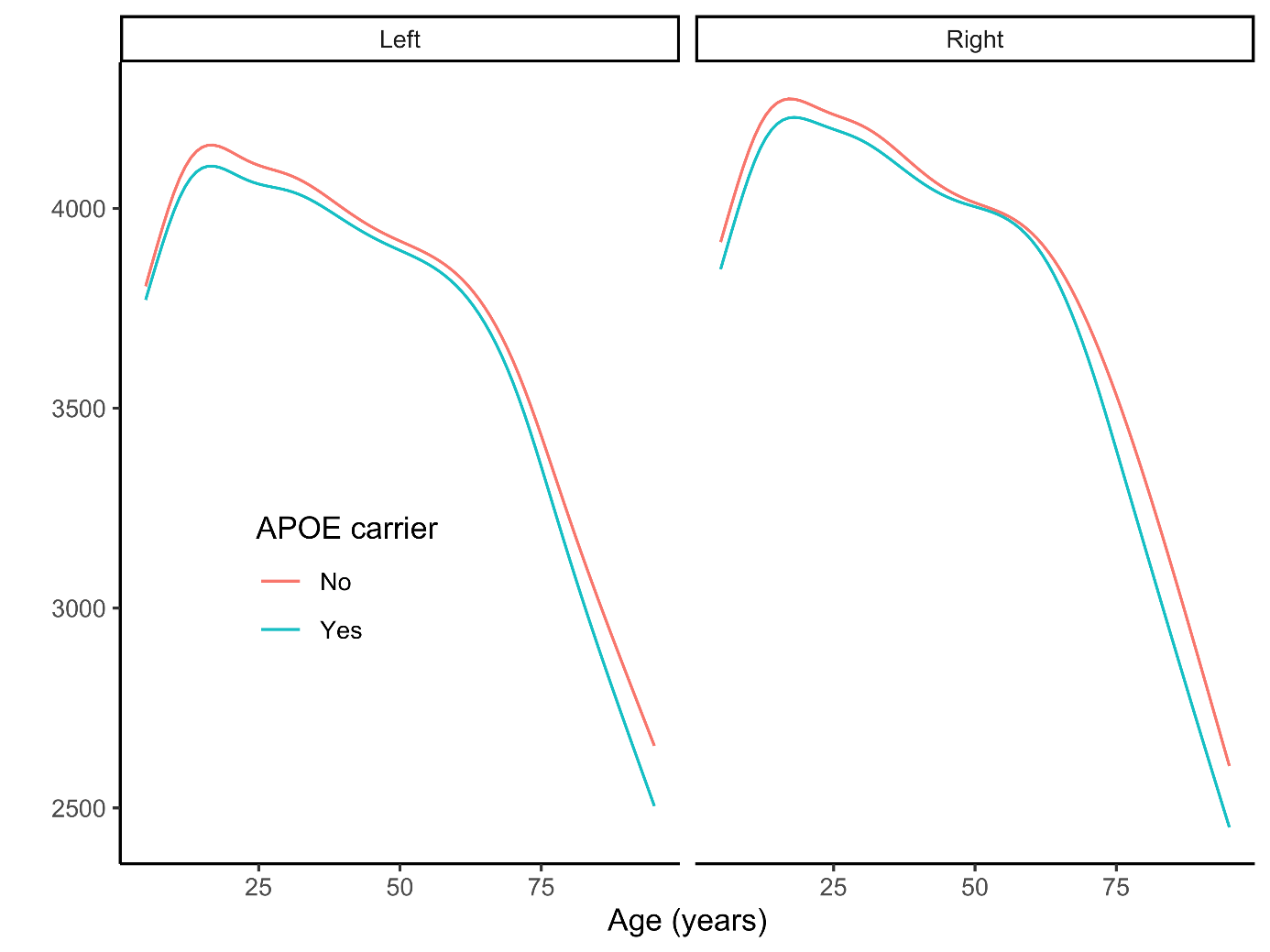
29. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2018.

30. Wood R. Generalized Additive Models: An Introduction with R: Chapman and Hall/CRC, 2006.

**Figure 1**

****

**Figure 2**

****