Elevated blood pressure accelerates white matter brain aging among late middle-aged women: a Mendelian Randomization study in the UK Biobank

Supplementary Materials

1. Supplementary Methods

1.1 Exclusion criteria

We excluded individuals with discordant self-reporting sex and genotype-inferred sex and individuals having extreme values of total volume of white matter hyperintensities (WMH) (those with total volume of WMH above Q3 + 1.5*IQR from T1 and T2_FLAIR images) to minimize the bias[1]. We also excluded those individuals who have taken anti-hypertensive medicine because their observed BP might not reflect the genetically predicted BP, thus introducing bias into the MR estimates[2-4]. To avoid selection bias towards people with relatively lower BP, we also performed a sensitivity analysis by including individuals taking antihypertensive treatment and redid the association analysis (see results in Table S9). The final pool includes N=228,473 individuals for analysis. There are a few self-reported diseases that may or may not confound our findings (including (non-)insulin-dependent diabetes mellitus, hypertensive heart disease, chronic ischemic heart disease, and other medical issues, see Table S2 for a complete list and the frequency counts), we decided to include these individuals to ensure we have sufficient sample size for GWAS and MR analysis.

1.2 QC of genotype data

We performed QC of genotype data and only kept the genetic variants with: minor allele frequency (MAF) ≥ 0.01, imputation quality score (INFO) > 0.1, Hardy-Weinberg equilibrium

exact test p-value (HWE) ≥ 0.001, missing genotype rate (GENO) ≤ 0.05 and missingness per individual (MIND) ≤ 0.2.

1.3 Definitions of covariates

Sex: sex was used phenotype code 31 (women and men) in the UKB.

Age: age was used phenotype code 21022 (Age at recruitment) in the UKB.

BMI: BMI was used phenotype code 21001 (Body mass index) in the UKB.

Alcohol consumption: alcohol drinker status was defined as current/past/never alcohol drinkers using 20117 (Alcohol drinker status) in the UKB.

Smoking status: smoking status was defined as current/past/never smokers using phenotype code 20116 (Smoking status)

Fruit consumption: fruit consumption was calculated by summing pieces of fresh fruit intake per day and pieces of dried fruit intake per day using phenotype codes 1309 (Fresh fruit intake) and 1319 (Dried fruit intake) in UKB.

Vegetable consumption: vegetable consumption was calculated by summing number of heaped tablespoons of cooked vegetables intake per day and number of heaped tablespoons of salad or raw vegetables intake per day using phenotype codes 1289 (Cooked vegetable intake) and 1299 (Salad/raw vegetable intake) in UKB.

Sedentary lifestyle: sedentary lifestyle was calculated by summing in a typical day, number of hours participants spend driving, using the computer (except for using the computer at work) and watching television using phenotype codes 1090 (Time spent driving), 1080 (Time spent using computer) and 1070 (Time spent watching television (TV)) in UKB.

1.4 Estimation of age bias corrected WM BAG using ML model

In the first part, we applied ML model to estimate the outcome WM BAG based on 39 regional FA measures and the chronological age, among those participants with both BP and FA data available. We used random forest (RF) regression to generate a function for estimating unbiased brain age in a training set of participants with no hypertension (Non-HTN Training in Fig 1; N=7,728). The parameters of the RF regression were tuned based on the coefficients of determination (R^2) between the chronological age and estimated brain age and mean absolute error (MAE) criteria to achieve the optimum predictive performance using a 5-fold cross-validation (CV). The RF regression was also used to select a set of FA features that have the most significant impact on brain aging. The locked ML model was then applied to the testing samples (Non-HTN and HTN Testing in Fig 1; N=7,728 and 1,445 respectively) to predict the brain age. The WM BAG was calculated by subtracting individuals' chronological age from their predicted brain age. The age-dependent bias has been noted to distort clinical interpretation in many brain age prediction studies[5,6]. We further used a simple linear regression model[7] to remove brain age prediction bias from WM BAG and evaluated the performance of our correction method by the MAE.

1.5 Linear association analysis between BP and BAG

In the second part, we applied a multiple linear regression model to test for the association between BP and BAG in MR sample 2, controlling for the aforementioned confounders.

1.6 Two-sample MR analysis to evaluate the causal effect of BP on BAG

In the last part, we performed a two-sample MR analysis to evaluate the causal effects of BP on BAG treating candidate genetic variants as IVs. The MR sample 1 was used to perform gene-exposure association analysis and select candidate IVs. Detailed steps of IV selection can be

found in the Supplementary Material section 1.4. The MR sample 2 was used to perform gene-outcome association analysis. The estimates from the two MR samples were combined using a generalized inverse variance weighted (gen-IVW) method to evaluate the causal effects of BP on BAG. The gen-IVW method took the ratio of gene-outcome association and gene-exposure association estimates and combined multiple dependent IVs into an overall estimate to assess the causal effect of exposure on the outcome while controlling the impact from LD between pairs of genetic variants[8,9]. Other popular MR methods such as MR-PRESSO [10] and MR-MIX [11] were also applied. For both MR analysis and association analysis, the analyses were performed in the general population as well as stratified by sex and age groups.

1.7 IV selection

We followed the three IV assumptions of MR analysis to select IVs:

(i) The IV is associated with the exposure;

(ii) The IV is independent of the confounding factors;

(iii) The IV is independent of the outcome given the exposure (i.e., the IV does not exert horizontal pleiotropy).

We first performed GWAS on BP using an additive genetic model adjusted for sex, age, BMI, genotyping chip type, and top 10 principle components (PCs) of population admixture in MR sample 1 to select genetic variants associated with blood pressure ($p<5 \times 10^{-8}$). We then performed a linkage disequilibrium (LD) clumping to remove genetic variants with $r^2>0.50$ within a 1000-kb window using Plink (version 1.9, www.cog-genomics.org/plink/1.9/)[12]. The parameter setting follows from recommendation in recent studies to balance between being too stringent (too small r^2, too wide window size, very few IVs selected) and too conservative (too large r^2, too narrow window size, too many IVs selected)[13-17]. To strengthen the IV selection

step, we included the BP GWAS results from the meta-analyzed ICBP cohort (N=757,601; partially overlapped with UKB cohort)[18] and only selected those IVs that passed the p-value thresholds ($p<5 \times 10^{-8}$) in both cohorts. Our MR method based on gen-IVW can handle dependent IVs while improving the strength of IVs[9]. Next, we removed IVs associated with any of the aforementioned confounders (Benjamini-Hochberg (BH) adjusted p-value > 0.05). Lastly, we performed conditional independence tests to exclude IVs with evidence of horizontal pleiotropy (BH adjusted p-value > 0.1) in MR sample 2. Additionally, we followed an MR guidance for IV selections proposed by Burgess et al.[19] to pick up IVs with a biological link to BP (i.e., SBP/DBP) listed in the NIH National Human Genome Research Institute GWAS catalogue (http://www.genome.gov/gwastudies) and annotated them using the online annotation tool FAVOR (http://favor.genohub.org/)[20]. These criteria substantially enhance the creditability of the selected IVs and the causal role of risk factor on outcome. The number of variants passing each IV selection step is summarized in Table S7.

1.8 Sensitivity analysis

To ensure our analysis are robust and unbiased and validate our major findings, we performed a series of sensitivity analyses as follows:

1. The current results used the continuous BP as the exposure. For robustness, we also performed analysis when treating blood pressure as binary variable using three different definitions of high blood pressure (diagnosed, stage-2 high blood pressure, and combined) and tested them in the BP/BAG linear association analysis. BP as binary variable 1: define high blood pressure as diagnosed only (International Classification of Diseases edition 10 (ICD-10) codes I10-I15 available in UKB). BP as binary variable 2: define high blood pressure as stage-2 high blood

pressure SBP >139/DBP >89. BP as binary variable 3: define high blood pressure as a combined diagnosed and stage-2 high blood pressure. The results remained largely consistent (Table S8).

2. We excluded the group of participants who took the anti-hypertensive medicine (exclusion criteria) in our main analysis for better genetically predicted BP in MR analysis. To reduce the selection bias, we also performed BP/BAG association analysis by including the group of participants who took the anti-hypertension medicine. The results remained largely consistent (Table S9).

3. For late middle-aged women, menopause and hormone level could be critical factors that impact the entire body including blood pressure and brain aging. Thus, we also further adjusted menopause (Yes/No) and hormone replacement therapy (Yes/No) as potential confounders in the subgroup analysis of the women aged 50-59 and the results were shown in Table S10. The results for this subgroup analysis remained significant.

4. To ensure the robustness of our main MR method gen-IVW, we additionally applied other popular MR methods for the same analysis, including MR-PRESSO[10] which corrects for horizontal pleiotropic outliers and MR-MIX[11] which adjusts genetic correlations. The results were shown in Table S11. In addition, a leave-one-out approach[21] was also used to evaluate the robustness of gen-IVW results (Table S11). The results were largely consistent.

6. Lastly, although the reverse causality is already mitigated in MR analysis since the genotypes are not generally susceptible to reverse causation and confounding[22], we performed another MR analysis by switching the exposure and the outcome to rule out the reverse causality (i.e. treating BAG as exposure and BP as outcome). The results were insignificant so we can rule out the possibility of reverse causality (Table S12).

2. Supplementary Results

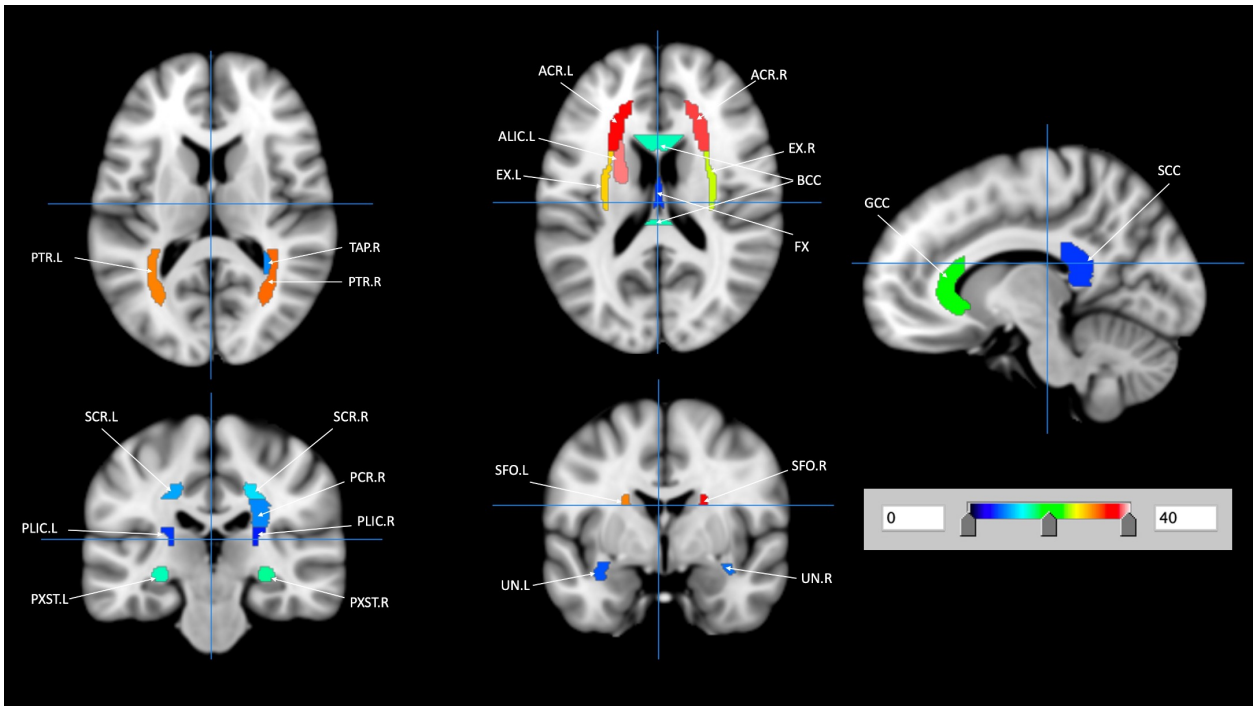

Figure S1. A total of 25 FA features were selected from the random forest model for the prediction of WM BAG. The color scale represents the negative natural log of minimum p-values of association coefficients between BP (i.e., SBP/DBP) and WM BAG.

Table S1-S12 can be found in the Excel File.

3. Reference

1. Jones DK, Lythgoe D, Horsfield MA, Simmons A, Williams SC, Markus HS. Characterization of white matter damage in ischemic leukoaraiosis with diffusion tensor MRI. Stroke 1999; 30:393-397.
2. Malik R, Georgakis MK, Vujkovic M, Damrauer SM, Elliott P, Karhunen V, et al. Relationship Between Blood Pressure and Incident Cardiovascular Disease: Linear and Nonlinear Mendelian Randomization Analyses. Hypertension 2021; 77:2004-2013.
3. Burgess S, Davies NM, Thompson SG, Consortium EP-I. Instrumental variable analysis with a nonlinear exposure-outcome relationship. Epidemiology 2014; 25:877-885.
4. Sun YQ, Burgess S, Staley JR, Wood AM, Bell S, Kaptoge SK, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. BMJ 2019; 364:l1042.
5. Le TT, Kuplicki RT, McKinney BA, Yeh HW, Thompson WK, Paulus MP, et al. A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. Front Aging Neurosci 2018; 10:317.
6. Mo C, Wang J, Ye Z, Ke H, Liu S, Hatch K, et al. Evaluating the causal effect of tobacco smoking on white matter brain aging: a two-sample Mendelian randomization analysis in UK Biobank. Addiction 2023; 118:739-749.
7. Beheshti I, Nugent S, Potvin O, Duchesne S. Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme. Neuroimage Clin 2019; 24:102063.
8. Mo C, Wang J, Ye Z, Ke H, Liu S, Hatch K, et al. Evaluating the causal effect of tobacco smoking on white matter brain aging: a two-sample Mendelian randomization analysis in UK Biobank. Addiction 2022.
9. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med 2016; 35:1880-1906.
10. Verbanck M, Chen C-y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature genetics 2018; 50:693-698.
11. Qi G, Chatterjee N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. Nat Commun 2019; 10:1941.
12. 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.
13. Lawlor D.A. WK, Borges M.C., Palmer T., Hartwig F.P., Hemani G., et al. A Mendelian Randomization Dictionary: Useful definistions and descriptions for undertaking , understanding and interpreting Mendelian Randomization studies. 2018.
14. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun 2018; 9:224.
15. Linner RK, Biroli P, Kong E, Meddens FW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nature Genetics 2019; 51:245-+.
16. Chen G, Wang Q, Xue R, Liu X, Yu H. Examining the Causal Inference of Leptin and Soluble Plasma Leptin Receptor Levels on Schizophrenia: A Mendelian Randomization Study. Front Psychiatry 2021; 12:753224.

17. Hou L, Yu Y, Sun X, Liu X, Yu Y, Li H, et al. Causal mediation analysis with multiple causally non-ordered and ordered mediators based on summarized genetic data. Stat Methods Med Res 2022; 31:1263-1279.
18. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 2018; 50:1412-1425.
19. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. Wellcome Open Res 2019; 4:186.
20. Zhou H, Arapoglou T, Li X, Li Z, Zheng X, Moore J, et al. FAVOR: functional annotation of variants online resource and annotator for variation across the human genome. Nucleic Acids Research 2023; 51:D1300-D1311.
21. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. Epidemiology (Cambridge, Mass) 2017; 28:30.
22. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep 2017; 4:330-345.