BRIGHT

Biostatistics Research on Imaging • Genomics • High-throughput Technologies

UNIVERSITY of MARYLAND SCHOOL OF MEDICINE

Evaluating the causal effect of tobacco smoking on white matter brain aging: a two-sample Mendelian randomization analysis in UK Biobank

Chen Mo, Zhenyao Ye, Tianzhou Ma, Shuo Chen

Division of Biostatistics and Bioinformatics and Maryland Psychiatric Research Center, School of Medicine, University of Maryland

INTRODUCTION

Tobacco smoking is a known risk factor associated with the accelerated decline of brain structures and functions during aging. However, the causal effect of smoking on brain aging remains unclear. To close this knowledge gap, we performed Mendelian randomization (MR) analysis to estimate the causal effect of smoking on white matter brain aging using genetic and neuroimaging data.

In the current study, we aimed to investigate the causal effects of smoking behaviors (i.e., smoking status (SS) and cigarette per day (CPD)) on Brain Age Gap (BAG) using the UK Biobank (UKB) cohort (1). Our results showed evidence of causal effect of smoking behaviors on BAG. These behaviors, such as being a smoker or smoking a larger number of cigarettes per day, could significantly cause impairment in brain microstructures reflected by an increase in BAG among the participants in UKB.

MATERIALS AND METHODS

Study Cohorts

Sample 1 UKB participants (N=12,907) who had both neuroimaging and smoking behavioral data. These participants had BAG data (see Outcome and BAG estimation for calculation details).

Sample 2(A) UKB participants who had smoking behavioral data yet no neuroimaging data (N=185,972). We calculated the summary statistics of genetic-exposure association for SS and CPD, separately.

Sample 2(B) We further used GSCAN summary statistics for CPD MR analysis to validate the MR results for CPD (2). The GSCAN cohort used had no UKB participants overlapped. The summary statistics are not available for SS in GSCAN.

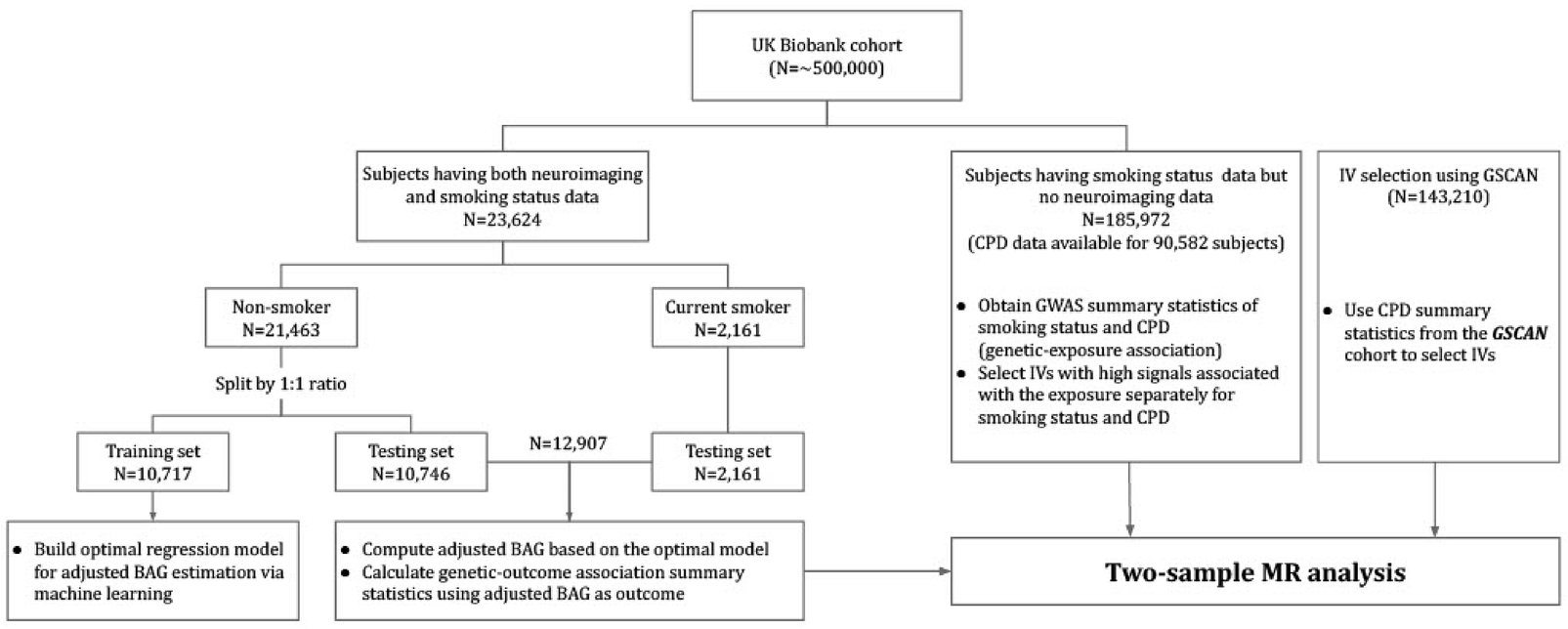

Figure 1. The flow diagram demonstrates the numbers of participants included in the two-sample Mendelian randomization analysis. We performed a two-sample Mendelian randomization (MR) analysis including two parts: instrumental variable (IV) selection (right) and outcome brain aging gap (BAG) analysis (left).

Exposures

Smoking behaviors (i.e., smoking status (SS) and cigarette per day (CPD)) were the exposures in our study.

Outcome and BAG Estimation

The outcome in our study was BAG. We used an independent set of UKB participants with both neuroimaging and smoking behavioral data (N=10,707) to train an optimal predictive model for calculating BAG for UKB participants in Sample 1 (N=12,907), following procedures below:

1) Implemented ML analysis through an internal 5-fold CV to achieve the optimal predictive performance using random forest (RF) regression
2) Estimated the predicted brain age ($B\hat{}$) using this optimal predictive model
3) Obtained BAG (ΔG) by subtracting chronological age (Y) by the predicted brain age ($\Delta G = Y - B\hat{}$)
4) Correcting the systematic bias caused by regression property (prediction tends to overestimate brain age at low chronological age and underestimate brain age at high chronological age)
5) Calculated the adjusted predictive brain age ($\hat{B}_{adj}$) and the **adjusted BAG ($\Delta G_{adj} = Y - \hat{B}_{adj}$) - This is the outcome**

GWAS and Summary Statistics

Genetic-Exposure Using Sample 1, we performed Genome-Wide Association Study (GWAS) analysis for each exposure under an additive genetic model adjusting for sex, age, BMI, genotyping chip type (i.e., UKBL/UKB chip) and top 10 principal components (PCs) of population admixture using PLINK (version 1.9) [3].

Genetic-Outcome Using Sample 2(A), we calculated the genetic-outcome association summary statistics by performing univariate association between the outcome and each potential IVs. The Sample 2(B) Genetic-Outcome summary statistics for CPD was downloaded from GSCAN website (https://genome.psych.umn.edu/index.php/GSCAN).

Two-Sample MR Analysis

A two-sample MR analysis was performed separately for using SS and CPD as the exposure using MR method with generalized weighted model (gen-IVW) (3). A p-value of 0.05 indicates a significant causal relationship. We first performed linkage disequilibrium pruning and clumping to remove genetic variants with r2 > 0.50). We then selected genetic variants as the instrumental variables (IVs) according to three criteria:

i. Genetic-exposure association p-value < 5×10-8
ii. Genetic-outcome p-value adjusted by the Benjamini-Hochberg false discovery rate method (BHp) > 0.05
iii. Genetic-confounder BHp > 0.05; the observed confounders included hypertension, diabetes, and stroke (as shown in Table 1)

These criteria were based on the IV assumptions in the MR analysis. The assumptions are visually shown in Figure 2 with more details in the legend.

Correspondence: *Chen Mo: cmo@bidmc.harvard.edu ; Shuo Chen: shuochen@som.umaryland.edu*

RESULTS

We selected 248 and 54 genetic variants as IVs in the MR models for SS and CPD, respectively. These IVs were well aligned with findings in existing studies, such as the gene-cluster CHRNA5-CHRNA3-CHRNB4 linked with CPD.

As shown in Table 1 and Figure 2B, the gen-IVW (UKB) (our primary method) showed a significant causal effect of SS on BAG ($\theta\hat{} = 0.21$; CI = 6.5×10-3, 0.41; p-value = 0.04). Also, CPD had significant causal effect on BAG (UKB: $\theta\hat{} = 0.16$; CI = 0.06, 0.26; p-value = 1.3×10-3; GSCAN: $\theta\hat{} = 0.16$; CI=0.18, 0.31; p-value = 0.03), indicating that the causal effect of CPD were robust across cohorts (see gen-IVW (GSCAN) in Figure 2B). The results showed that smokers were 0.21 years older in brain age than comparable non-smokers at the same chronological age, and having an extra cigarette per day increased brain age by additional 0.16 years.

We also performed sensitivity analyses using some commonly used MR methods, including MR-weighted-median, MR-PRESSO, MR-MIX, and contamination mixture method (MRconmix), MR-Egger. These analyses showed robust causal effects of SS and CPD separately on BAG (see Figure 3B). The I2 statistics suggested that MR-Egger had sufficiently low reliability for CPD (I2 = 0.0%). For SS, I2 = 52% also suggested violation of the NOME assumption in MR-Egger. Therefore, the results were excluded. Additionally, we performed MR analysis to examine the other possible causal direction using BAG as the exposure and smoking behaviors (SS and CPD) as the outcome. We found no evidence of reverse causation between BAG and smoking behaviors as MR *p*-values were larger than 0.05

Table 1. Results of Mendelian randomization and association analyses in the study sample.

	Association	
	$\hat{\beta}$ ± SE (95% CI)	p-value
SS	0.88 ± 0.07 (0.74, 1.02)	2.81×10^{-36}
CPD	0.05 ± 0.01 (0.03, 0.07)	5.49×10^{-6}
	Mendelian randomization	
	$\hat{\theta}$ ± SE (95% CI)	p-value
SS	0.21 ± 0.10 (6.5×10^{-3}, 0.41)	0.04
CPD	0.16 ± 0.05 (0.06, 0.26)	1.30×10^{-3}

SE = standard error; CI = confidence interval;
SS = smoking status; CPD = cigarette per day.

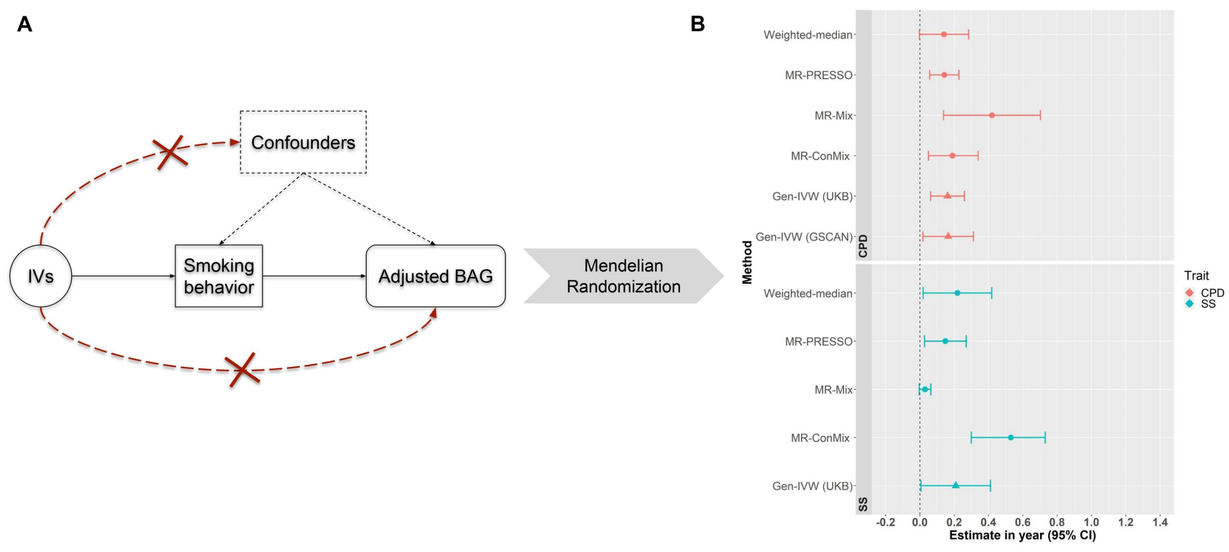

Figure 2. Mendelian randomization and the results. A) visually demonstrates the three fundamental IV assumptions in the MR analysis: I) IVs are significantly associated with the exposure (i.e., SS or CPD); II) the exposure is not significantly associated with confounders of the exposure-outcome association, and III) IVs can affect the outcome variable only through the exposure. B) displays the causal effect estimate with a 95% confidence interval using different MR methods for SS and CPD separately, based on the IVs selected following the three IV assumptions. Gen-IVW (marked with a triangle) is the primary MR method (i.e., weighted generalized linear regression), and the other methods (marked with a dot) are the MR methods used in the sensitivity analysis.

FINAL REMARKS

- Our study provided a thorough analysis of causal inference for smoking behaviors on BAG and revealed robust and consistent causal effects in a large study cohort. The estimated causal effect can reveal the overall effect of smoking and its biological consequences (e.g., cardiovascular condition worsening) on brain aging and thus unfolds the reverse effect on brain aging by quitting tobacco smoking.
- Our findings confirmed the causal effect of smoking on accelerated neural degeneration during aging interpreted by age year.
- The significant causal effect of smoking on BAG suggested that smoking prevention can be an effective intervention for accelerated brain aging and the age-related cognitive function decline.

SELECTED REFERENCES

1 Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015; 12:e1001779.

2 Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51(2):237–44.

3 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(11):1880–906.