Evaluation of the Causal Relationship Between Smoking and Schizophrenia in Asia

Mei-Hsin Su\textsuperscript{a,b}, Rou-Yi Lai\textsuperscript{b}, Yen-Feng Lin\textsuperscript{c}, Chia-Yen Chen\textsuperscript{d,e}, Yen-Chen A. Feng\textsuperscript{f}, Po-Chang Hsiao\textsuperscript{f}, Shi-Heng Wang\textsuperscript{a,b,*}

\textsuperscript{a}Department of Occupational Safety and Health, College of Public Health, China Medical University, Taichung, Taiwan
\textsuperscript{b}Department of Public Health, College of Public Health, China Medical University, Taichung, Taiwan
\textsuperscript{c}Center for Neuropsychiatric Research, National Health Research Institutes, Miaoli, Taiwan;
\textsuperscript{d}Biogen, Cambridge, MA, USA;
\textsuperscript{e}Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
\textsuperscript{f}College of Public Health, National Taiwan University, Taipei, Taiwan

\textsuperscript{*} Corresponding author. College of Public Health, China Medical University, Taichung, Taiwan.

Email address: wangsh@mail.cmu.edu.tw (S.H. Wang)
Abstract
Cigarette smoking has been suggested to be associated with the risk of schizophrenia (SCZ) in observational studies. A significant causal effect of smoking on SCZ has been reported in the European population using the Mendelian randomization (MR) approach; however, no evidence of causality was found in participants from East Asia (EAS). Using the Taiwan Biobank (TWBB, sample size up to 79,989), we conducted genome-wide association studies (GWAS) to identify susceptibility loci for smoking behavior, which included the initiation of smoking and the onset age. To maximize the power of genetic discovery in the EAS population, we meta-analyzed GWAS from the TWBB and Biobank Japan (BBJ, sample size up to 165,436) for smoking traits. The GWAS for SCZ was taken from the Asia Psychiatric Genomics Consortium, which included 22,778 cases and 35,362 controls. We performed a two-sample MR to estimate the causality of smoking behavior on SCZ in the EAS population. In TWBB, we identified one novel locus that met genome-wide significance for onset age. In a meta-analysis of TWBB and BBJ, we identified two novel loci for smoking initiation. In MR, a marginal significance was found for the causality of smoking initiation on SCZ (odds ratio (OR) = 4.00, 95% confidence interval (CI) = 0.89–18.01, P = 0.071). Later onset age for smoking was causally associated with a lower risk of SCZ (OR for a per-year increase in onset = 0.96, 95% CI = 0.91–1.01) with a marginal significance (P = 0.098).

Keywords: Mendelian randomization, smoking, schizophrenia, genome-wide association study
1. Introduction

Schizophrenia (SCZ) is often comorbid with smoking behaviors. Both case-control and cohort studies indicated that smokers were at a significantly higher risk of SCZ than non-smokers (Weiser et al., 2004), especially heavy-smokers (Hunter et al., 2020). Meta-analyses, including cohort studies, also indicated that smokers have a two-fold increased relative risk of SCZ than non-smokers (Gurillo et al., 2015; Hunter et al., 2020). In addition, the comorbidity of smoking and SCZ is associated with poorer prognosis (Whiteford et al., 2013) and aggravated mental symptoms, such as flat affect, delusions, and hallucinations (Kelly & McCreadie, 1999).

Some possible explanations have been proposed for the association between smoking behaviors and SCZ. One argument is that substance use may lead to SCZ. According to reports, 90% of patients with SCZ started to smoke before the onset of their illness (Kelly & McCreadie, 1999), and adolescents with smoking behaviors were more likely to develop SCZ than non-smoking adolescents (de Leon, 1996; Weiser et al., 2004). These studies used temporal sequence approach to infer the possibility that substance use may lead to SCZ. Another possible reason is shared genetic architecture. In our previous study, we calculated the polygenetic risk scores of SCZ based on genome-wide association studies (GWAS), and found that the score was positively associated with lifetime tobacco smoking (Wang et al., 2019). On the other hand, the polygenic risk scores of cotinine concentrations also significantly
predict schizophrenia diagnosis (Chen et al., 2016). In addition, a genetic correlation was found between SCZ and regular smoking (Harper et al., 2021). Self-medication is another hypothesis for the association between smoking and SCZ. Patients with SCZ may try to diminish their symptoms and the side effects of antipsychotic drugs through nicotine consumption, which can increase hepatic clearance and restore the dopamine blockade (Winterer, 2010). Animal models also provided evidence that chronic nicotine administration reversed hypofrontality by nicotinic acetylcholine receptor (nAChR) modulation and further attenuated the symptoms of SCZ (Koukouli et al., 2017; Noda et al., 2020).

The association between smoking and SCZ has been established, but the causal relationship between them remains conflicting. Investigating the causal relationship between smoking and SCZ may help in blocking the path to SCZ and subsequently reduce the disease burden. The Mendelian randomization (MR) is a novel study design for investigating causality using a genetic approach. Reports suggest that smoking plays a causal role in SCZ (Barkhuizen et al., 2021; Gage et al., 2017; Wootton et al., 2020; Yuan et al., 2020). Further, the causal effect of SCZ on smoking initiation behaviors was also detected (Wootton et al., 2020). However, most studies were conducted in Caucasian populations with a sufficient sample size for GWASs for smoking and SCZ. Only one MR analysis from the Asian population did not support
such causality (Chen et al., 2021). The negative finding may have resulted from insufficient power for the MR analysis with a limited sample size for GWAS. Therefore, estimating the causal relationship with a larger sample size in the Asian population can help provide precise results.

Since smoking behaviors are associated with various physical (Moudi et al., 2018; Teramoto et al., 2021) and mental illnesses (Hunter et al., 2020; Wang et al., 2019), investigating the causality between them helps block the pathway which may lead to SCZ. Therefore, this study aimed to conduct a well-powered GWAS for smoking initiation and onset age among East Asian populations with larger sample size which came from meta-analyses of the Taiwan Biobank (TWBB) and Biobank Japan (BBJ). We then examined the causal relationship between these two smoking behaviors and SCZ in the Asian population.

2. Method

2.1. GWAS for smoking

This study used individual genotyping and phenotyping data from the TWBB, the largest government-supported biobank in Taiwan since 2012. The TWBB recruits community-based samples aged 30–70 years who are cancer-free at recruitment. The internal review board approved the recruitment and data collection procedures of TWBB. Each participant signed an approved informed consent form, provided blood
samples, and underwent physical examinations and face-to-face interviews. This study was approved by the Central Regional Research Ethics Committee of China Medical University, Taichung, Taiwan (CRREC-108-30).

Genotyping of 95,238 TWBB participants was performed using customized TWBB chips and processed on the Axiom Genome-Wide Array Plate System (Affymetrix, Santa Clara, CA, USA). Furthermore, 26,274 participants were genotyped on the TWBv1 chip, and 68,964 participants were genotyped on the TWBv2 chip. We conducted quality control and imputation of the two chips separately. Quality control included the exclusion criteria of variants with call rate < 95%, individuals with more than 5% missing variants, minor allele frequency (MAF) < 0.001, and deviation from Hardy-Weinberg equilibrium with $P < 1 \times 10^{-6}$. Imputation was performed based on the 973 TWBB panels from whole-genome sequencing in TWBB participants and 504 EAS panels in the 1000 Genomes project, variants with MAF $\geq 0.5\%$, and imputation INFO score $\geq 0.7$. In addition, cryptic relatedness was removed, and we estimated identity by descent (IBD) sharing coefficients, PI-HAT = probability (IBD = 2) + 0.5 x probability (IBD = 1), between any two participants in KING and excluded one individual from a pair with PI-HAT $\geq 0.1875$.

GWAS for two smoking behaviors included smoking initiation (yes/no) and onset
age in a sample size of 79,989 individuals. We performed linear or logistic regression in PLINK for association tests with adjustment for age, age$^2$, sex, age by sex interaction, age$^2$ by sex interaction, and the top 20 principal components. In addition, we performed a genetic association test separately for the two chips and performed an inverse-variance-weighted fixed-effect meta-analysis in METAL. For each trait, the sample size and heritability estimates for each chip and genetic correlation estimates (Bulik-Sullivan et al., 2015a, 2015b) between the two chips are detailed in Supplementary Table 1.

To maximize the power of genetic discovery in East Asia, we meta-analyzed GWAS from the TWBB and BBJ (sample size up to 165,436, for each smoking trait with MAF $\geq$ 0.5% and INFO score $\geq$ 0.7) (Matoba et al., 2019), including smoking initiation (N = 245,425) and onset age (N = 46,000). The genetic correlation between TWBB and BBJ was median to high for smoking initiation but not for onset age (Supplementary Table 2).

2.2. GWAS summary for SCZ

The GWAS summary of SCZ from the Asia Psychiatric Genomics Consortium (PGC), including 22,778 cases and 35,362 controls (Lam et al., 2019) from East Asia, and the diagnosis of SCZ was based on the Diagnostic Manual of Mental
Disorders-IV (DSM-IV) system.

### 2.3. Mendelian randomization

**Genetic instrument for causality of smoking on SCZ**

We mapped the variants from exposure (smoking phenotypes) data to outcome (SCZ) GWAS and preserved those that could be mapped to both. Linkage disequilibrium (LD) clumping was conducted based on $r^2 > 0.0001$ within a 1,000 kb window to select independent variants. To recruit enough instrumental variables (at least five), we used a lenient threshold for single nucleotide polymorphism (SNP) selection and identified eight SNPs for smoking initiation with a $P$-value of $5 \times 10^{-7}$ and five SNPs for onset age with a $P$-value of $1 \times 10^{-6}$.

**Genetic instruments for bidirectional MR**

A bidirectional MR was conducted to estimate the causal effect of SCZ on smoking behaviors and examine whether an MR study supports the argument of self-medication in patients with SCZ. The data for bidirectional MR were taken from GWAS summary data of the Asian population with 18 independent SNPs of genome-wide significance ($P < 5 \times 10^{-8}$) for both smoking initiation and onset age.

**MR analysis**
In this study, five MR methods were applied as follows: inverse variance-weighted (IVW) (Burgess et al., 2013; Lawlor et al., 2008), weighted median (Bowden et al., 2016), weighted mode (Hartwig et al., 2017), Mendelian Randomization Pleiotropy Residual Sum and Outlier (MR-PRESSO) (Verbanck et al., 2018), and MR Egger (Bowden et al., 2015). IVW provides the best unbiased estimation in the absence of pleiotropy and under valid SNPs; hence, the main results for MR analysis are based on IVW method and the other four MR methods for the sensitivity analyses. We also examined pleiotropy using Egger intercept and heterogeneity using Cochran’s Q and Rücker’s Q.

2.4. Genetic correlation

Genetic correlation between two smoking traits and SCZ among East Asians was conducted to detect whether common genetic variants existed between these two phenotypes. The genetic correlation was measured using LD score regression (Bulik-Sullivan et al., 2015a, 2015b).

3. Results

3.1. Smoking GWAS in the Asian population

The Manhattan plot for two smoking traits for TWBB is shown in Figure 1. One novel locus was identified in TWBB (rs553874586) (Table 1). Rs553874586 in the
**MMD** gene on chromosome 17 was associated with onset age (beta = -3.77, SE = 0.654, \(P = 8.08 \times 10^{-9}\)). In the GWAS meta-analysis of TWBB and BBJ, three SNPs met genome-wide significance, and all of them were novel (Table 1). Among these novel loci, variants rs79574881 at 8p22 (beta = -0.02, SE = 0.004, \(P = 2.84 \times 10^{-8}\)) and rs2156008 at 18q21.2 (beta = 0.01, SE = 0.002, \(P = 3.21 \times 10^{-8}\)) are associated with smoking initiation, and rs553874586 at 17q22 (beta = -3.77, SE = 0.654, \(P = 8.07 \times 10^{-9}\)) is associated with onset age.

### 3.2. Causal relationships of smoking and SCZ by MR study

Scatter plots of the causal estimation for smoking behaviors on SCZ is presented in Supplementary Figure 1, and forest plots are presented in Supplementary Figure 2. No heterogeneity was detected in the MR analysis for causality of the two smoking traits on SCZ (Supplementary Table 3).

The IVW method showed that smoking initiation was causally associated with an increased risk of SCZ with marginal significance (odds ratio [OR] = 4.00, 95% confidence interval [CI] = 0.89–18.01, \(P = 0.071\)) (Table 2). MR-PRESSO detected no outliers and no overall pleiotropy (\(P\)-value for MR Egger intercept = 0.99). Other MR methods provided similar estimates but did not reach statistical significance.

For causality of onset age of smoking on SCZ, the IVW method showed a marginal significance. A later onset age for smoking was causally associated with a
lower risk of SCZ (OR for a per-year increase in onset = 0.96, 95% CI = 0.91–1.01, \( P = 0.098 \)). There was no overall pleiotropy (\( P \)-value for MR Egger intercept = 0.68) (Table 2).

### 3.3. Bidirectional MR

We used Asia PGC data to identify genetic instruments for SCZ. The MR analyses did not support the causality of SCZ on the two smoking traits (see Supplementary Table 4).

### 3.4. Genetic correlation of smoking and SCZ

No significant genetic correlation was found between smoking behavior and SCZ among the Asian population. The genetic correlations were low, both for smoking initiation (\( r_g = -0.003, P = 0.94 \)) and onset age (\( r_g = 0.102, P = 0.17 \)) and SCZ.

### 4. Discussion

Using improved-power GWAS data from the EAS population, we detected three novel variants related to smoking initiation and onset age. The MR analysis provided evidence for the harmful effect of smoking initiation and earlier onset age on SCZ in the EAS population, which was in line with previous evidence in the European population.
Three novel variants were detected in relation to smoking behaviors in a meta-analysis of the Asian population. Among these SNPs, rs79574881 and rs2156008 were associated with smoking initiation. Rs79574881 is located on the DLC1 gene, which encodes a GTPase-activating protein. It is a tumor suppressor gene and has been reported to be related to prostate, lung, and breast cancers (Tripathi et al., 2014; Wang et al., 2020). Further, variants in the DLC1 gene have also been associated with nicotine dependence (Gelernter et al., 2015). Rs2156008 is located on the TCF4 gene, which is a protein-coding gene related to nervous system development. Many SNPs in the TCF4 gene are associated with SCZ or SCZ endophenotypes (Quednow et al., 2014). Moreover, the gene effects were modulated by smoking behavior, where heavy-smokers showed stronger gene effects on the SCZ endophenotype than light-smokers and non-smokers (Quednow et al., 2012).

Rs553874586 is a novel SNP associated with the onset age of smoking in our study. This SNP is located on the MMD gene, which is a protein-coding gene, and its molecular functions include protein kinase activity and signaling receptor activity.

Consistent and strong evidence has been reported for the causal effect of smoking initiation on SCZ among European-origin individuals (Barkhuizen et al., 2021; Gage et al., 2017; Wootton et al., 2020; Yuan et al., 2020); however, this issue has been less discussed in the Asian populations. Only one MR analysis in Asian participants with
limited statistical power reported no significant causal association between smoking and SCZ (Chen et al., 2021). Concerning that the biological mechanism should be presented in different populations if the causal effect between two phenotypes is truly existence, we increased the sample size for exposure and outcome variables and found that smoking initiation showed a marginal causal association with SCZ. Hence, partial evidence was found for the causal effect of smoking initiation on SCZ among Asian populations for a sufficient sample size, which corroborates with previous findings in the European population. In addition to previous MR studies, our results are supported by studies of various study designs. One retrospective study reported that 90% of patients with SCZ started smoking before the onset of their disease (Kelly & McCreadie, 1999), and longitudinal studies have pointed out that cigarette smoking increases the risk of developing SCZ and consistently reported a dose-response relationship of smoking quantity (Kendler et al., 2015; McGrath et al., 2016; Mustonen et al., 2018; Weiser et al., 2004). A previous MR study with 632,802 Caucasian population of smoking initiation GWAS also supported our finding, which indicated a causal role of smoking initiation in many psychiatric disorders, including SCZ, depression, and bipolar disorder (Barkhuizen et al., 2021).

Previous studies have provided evidence of shared genetic variants between smoking and SCZ. For example, researchers found that the risk of SCZ increased with
an increase in the polygenic risk score of plasma cotinine concentration, and a higher genetic score for SCZ increased the risk of smoking behaviors (Chen et al., 2016). Another study also reported a positive correlation between polygenic risk scores between regular smoking and SCZ (Harper et al., 2021). This implies some shared genetic liability between SCZ and smoking behaviors, which also describes the possible biological mechanism between smoking behaviors and SCZ. However, the genetic correlation between these two phenotypes did not achieve statistical significance in our study, which may have resulted from LD varying across different populations (Hill, 2013).

Brain structure may explain the causality of smoking and SCZ. For example, researchers have observed that event-related electroencephalogram activity (P3/δ component) showed a negative correlation with SCZ polygenic risk scores and a prefrontal control component (θ/antisaccade) was negatively correlated with number of drinks per week and regular smoking polygenic risk scores (Harper et al., 2021). Because the brain structure and function in different brain regions are closely related, this finding provides evidence that brain structure may simultaneously affect the SCZ status and smoking behaviors.

Using a genetic approach to investigate the causality between smoking behaviors and SCZ can avoid possible confounders and ensure the temporal sequence. Although
investigating an Asian population provided an opportunity to investigate causality particularly in this population, the sample size for smoking GWAS was insufficient to provide enough instruments that met genome-wide significance \((P = 5 \times 10^{-8})\). Instead, we used a wider P-value threshold at \(P = 5 \times 10^{-7}\) for smoking initiation and \(P = 1 \times 10^{-6}\) for onset age to include sufficient SNPs. This may recruit weak instrumental variables and lead to bias in MR analysis. Furthermore, we only included two smoking behaviors (smoking initiation and onset age of smoking) in the analysis; hence, the causal effect of a wide range of smoking traits on SCZ needs to be examined further.

In conclusion, this study identified two novel SNPs related to smoking initiation and one novel SNP related to the onset age. We also provided evidence for the marginal causal effect of smoking initiation on SCZ, which corresponds to the longitudinal and cross-sectional studies of the association between smoking and SCZ (de Leon, 1996; Kelly & McCreadie, 1999; Weiser et al., 2004). SCZ is a complex trait, and MR analysis can help identify possible causal risk factors. Future efforts to elucidate the mechanisms underlying the association between smoking and SCZ are needed and may help early prevention.

**Acknowledgement**
This work was supported by the Taiwanese National Health Research Institutes (NHRI-EX109-10931PI, NHRI-EX110-10931PI, and NHRI-EX111-10931PI) and
China Medical University (CMU110-MF-79).

**Data availability**

Reference


Wootton, RE., Richmond, RC., Stuijfzand, BG., Lawn, RB., Sallis, HM., Taylor,

Figure Legend

Figure 1. Manhattan plot for smoking behaviors among Taiwan Biobank data, including (a) smoking initiation, (b) onset age. The red line indicates the genome-wide significant level ($p<5E-08$) and the blue line indicates suggest significant level ($p<1E-05$).

(a) Smoking initiation

(b) Onset age
Table 1. Genome-wide association analysis for smoking behavior in Asia population

<table>
<thead>
<tr>
<th>Lead SNP</th>
<th>chr</th>
<th>Position (hg38)</th>
<th>Effect allele</th>
<th>gene</th>
<th>TWBB</th>
<th>BBJ</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking initiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs79574881</td>
<td>8</td>
<td>13420362</td>
<td>A</td>
<td>DLC1</td>
<td>-0.06</td>
<td>0.046</td>
<td>0.2039</td>
</tr>
<tr>
<td>rs2156008</td>
<td>18</td>
<td>55614013</td>
<td>A</td>
<td>TCF4</td>
<td>0.06</td>
<td>0.014</td>
<td>7.90E-06</td>
</tr>
<tr>
<td>Onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs553874586</td>
<td>17</td>
<td>55405197</td>
<td>A</td>
<td>MMD</td>
<td>-3.77</td>
<td>0.654</td>
<td>8.08E-09</td>
</tr>
</tbody>
</table>

Beta SE p
2.84E-08
3.21E-08
8.07E-09
## Table 2. The causal estimation of smoking behaviors on schizophrenia in Asia population

<table>
<thead>
<tr>
<th>Exposure</th>
<th>N</th>
<th>MR Method</th>
<th>N outliers</th>
<th>OR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking initiation</td>
<td>8</td>
<td>IVW</td>
<td></td>
<td>4.00 (0.89-18.01)</td>
<td>0.0714</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted median</td>
<td></td>
<td>3.44 (0.56-21.11)</td>
<td>0.1826</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted mode</td>
<td></td>
<td>4.57 (0.30-69.27)</td>
<td>0.3091</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR-PRESSO</td>
<td>0</td>
<td>4.00 (1.78-9.98)</td>
<td>0.0122</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR Egger</td>
<td>(intercept, SE)</td>
<td>-0.0003 (0.0281)</td>
<td>0.9906</td>
</tr>
<tr>
<td>Onset</td>
<td>5</td>
<td>IVW</td>
<td></td>
<td>0.96 (0.91-1.01)</td>
<td>0.0982</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted median</td>
<td></td>
<td>0.97 (0.91-1.03)</td>
<td>0.5071</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted mode</td>
<td></td>
<td>0.97 (0.90-1.05)</td>
<td>0.3311</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR-PRESSO</td>
<td>0</td>
<td>0.96 (0.91-1.01)</td>
<td>0.1736</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MR Egger</td>
<td>(intercept, SE)</td>
<td>0.93 (0.83-1.05)</td>
<td>0.3496</td>
</tr>
</tbody>
</table>

Note: CC-BY-NC-ND 4.0 International license.