# Appraising the causal role of smoking in idiopathic pulmonary fibrosis: a Mendelian randomization study

Jiahao Zhu,[1] Dan Zhou,[2,3] Min Yu,[4] Yingjun Li[5]

[1]Department of Epidemiology and Health Statistics, Hangzhou Medical College, Hangzhou, Zhejiang, China
[2]Department of Big Data in Health Science, School of Public Health, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[3]Vanderbilt University Medical Center, Nashville, Tennessee, USA
[4]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, Zhejiang, China
[5]Department of Epidemiology and Health Statistics, School of Public Health, Hangzhou Medical College, Hangzhou, Zhejiang, China

**Correspondence to**
Prof. Min Yu, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, 310051, Zhejiang, China; mycdc1234@163.com and Dr Yingjun Li, Department of Epidemiology and Health Statistics, School of Public Health, Hangzhou Medical College, Hangzhou, 310053, Zhejiang, China; 2016034036@hmc.edu.cn
Published Online First 22 May 2023



► http://dx.doi.org/10.1136/thorax-2023-220483



**To cite:** Zhu J, Zhou D, Yu M, *et al*. *Thorax* 2024;**79**:179–181.

## ABSTRACT

Smoking has been considered a risk factor for idiopathic pulmonary fibrosis (IPF) in observational studies. To assess whether smoking plays a causal role in IPF, we performed a Mendelian randomization study using genetic association data of 10 382 cases with IPF and 968 080 controls. We found that genetic predisposition to smoking initiation (based on 378 variants) and lifetime smoking (based on 126 variants) were associated with a higher risk of IPF. Our study suggests a potential causal effect of smoking on increasing IPF risk from a genetic perspective.

## INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a lethal lung disease characterised by progressive fibrosis of lung parenchyma for which there is currently no cure.[1] Although the cause of IPF is unclear, tobacco smoking is thought to play a part in the pathogenetic processes of IPF.[1] However, whether smoking represents a causal determinant remains uncertain, because the available evidence is scarce and originates mainly from observational studies, which are vulnerable to confounding bias and reverse causation. Mendelian randomization (MR) is an increasingly used approach that overcomes these challenges by exploiting randomly allocated genetic variants as instruments to make reliable causal inferences. Here, we conducted a MR study to investigate the causal association between smoking and the risk of IPF.

## METHODS

In this study, we applied a two-sample design based on summary data from genome-wide association studies (GWASs). As genetic instruments, 378 independent, genome-wide significant ($p<5\times10^{-8}$) single-nucleotide polymorphisms (SNPs) associated with smoking initiation (ever vs never being a regular smoker) were identified from the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN), involving 1.2 million individuals.[2] In a secondary analysis, we used 126 independent SNPs associated with lifetime smoking (a continuous measure that takes into account smoking initiation, duration, heaviness, and cessation) as genetic instruments from the UK Biobank with 462 690 participants.[3] These instrumental SNPs explained 2.3% of the variance in smoking initiation and 1.3% in lifetime smoking and have been treated as robust instruments with *F*-statistics>10 in prior MR studies.[4] GWAS summary data for IPF were derived from a meta-analysis of 5 cohorts (UK, Chicago, Colorado, UUS [US, UK, and Spain], and Genentech Study) by the International IPF Genetics Consortium (4125 cases and 20 464 controls)[5] and another meta-analysis of 9 biobanks (BioVU, Colorado Centre for Personalised Medicine, Estonian Biobank, FinnGen, HUNT Study, Michigan Genomics Initiative, Mass General Brigham, UCLA Precision Health Biobank, and UK Biobank) by the Global Biobank Meta-analysis Initiative (6257 cases and 947 616 controls).[6] IPF was defined using European Respiratory Society/American Thoracic Society guidelines in the International IPF Genetics Consortium and using International Classification of Diseases codes (515 and 515.0 for ninth Revision and J84.1, J84.8, J84.89, J84.17, J84.1, and J84.10 for 10th Revision) in the Global Biobank Meta-analysis Initiative. GWAS models have adjusted for age, sex, and study-specific covariates where possible in the original studies. All participants included in the study were of European ancestry. The smoking and IPF studies involved some overlapping participants (43% of subjects in the Global Biobank Meta-analysis Initiative from the UK Biobank where the instruments for lifetime smoking were identified). However, sample overlap is not expected to introduce significant bias, because strong instruments (eg, *F*-statistics>10) for smoking were used and these overlapping samples were from large biobanks (eg, UK Biobank).

The principal analysis was performed using the inverse-variance weighted (IVW) method. Sensitivity analyses robust to pleiotropy were conducted, including weighted median, MR-pleiotropy residual sum and outlier (MR-PRESSO), and MR-Egger (summarised in table 1). The estimates from two IPF datasets were combined using random-effects meta-analysis. We examined horizontal pleiotropy and heterogeneity using the MR-Egger intercept and the Cochran's Q statistic, respectively. Reverse MR was performed to test for the potential reverse causation using 23 SNPs associated with IPF as genetic instruments.[5] Characteristics of instrumental SNPs are given in online supplemental tables 1–3. Statistical analyses were conducted using R (version 3.6.3) with "TwoSampleMR" and "MRPRESSO" packages. The significant threshold was 2-tailed p<0.05.

## RESULTS

The IVW results showed that genetic predisposition to smoking initiation (OR (OR) = 1.29; p=0.002) and lifetime smoking (OR=1.63; p<0.001) were associated with an increased risk of IPF in the meta-analysis of two datasets (table 2). The direction of effect was consistent across datasets, although the association in the International IPF Genetics Consortium did not reach the significant threshold. There were indications

**Table 1** Summary of applied Mr methods

| Method | Assumptions | Strengths | Weaknesses | PubMed ID |
|---|---|---|---|---|
| IVW | All genetic instruments are valid. | Has optimal statistical power. | Estimates are biased if there is directional pleiotropy. | 24114802 |
| Weighted median | More than 50% of the weight comes from valid genetic instruments. | Informs the estimate supported by the majority of evidence. | May be less efficient. | 27061298 |
| MR-PRESSO | The largest group of candidate instruments with similar estimates is the group of valid instruments. | Detects outliers and provides an estimate after removal of outliers | May have high false-positive rate. | 29686387 |
| MR-Egger | Associations of genetic instruments with the exposure are uncorrelated with any pleiotropic effects of the instruments on the outcome (InSIDE assumption). | Quantifies directional pleiotropy and provides a consistent estimate even if all genetic variants are pleiotropic. | Has low precision and can be strongly influenced by outliers. | 29040600 |

IVW, inverse variance weighted; MR, Mendelian randomization; PRESSO, pleiotropy residual sum and outlier.

**Table 2** Associations of genetic predisposition to smoking initiation and lifetime smoking with the risk of IPF

| | IVW (random effects) | | Weighted median | | MR-PRESSO | | MR-Egger | |
|---|---|---|---|---|---|---|---|---|
| | OR (95% CI) | *P*-value | OR (95% CI) | *P*-value | OR (95% CI) | *P*-value | OR (95% CI) | *P*-value |
| Smoking initiation* | | | | | | | | |
| International IPF Genetics Consortium | 1.17 (0.95, 1.43) | 0.137 | 1.08 (0.85, 1.29) | 0.523 | 1.12 (0.92, 1.36) | 0.253 | 1.50 (0.64, 3.48) | 0.349 |
| Global Biobank Meta-analysis Initiative | 1.37 (1.17, 1.60) | <0.001 | 1.30 (1.05, 1.59) | 0.014 | 1.37 (1.17, 1.60) | <0.001 | 1.77 (0.91, 3.47) | 0.096 |
| Meta-analysis (random effects) | 1.28 (1.10, 1.49) | 0.002 | 1.21 (1.01, 1.43) | 0.034 | 1.25 (1.03, 1.52) | 0.025 | 1.67 (0.98, 2.81) | 0.059 |
| Lifetime smoking† | | | | | | | | |
| International IPF Genetics Consortium | 1.37 (0.93, 2.02) | 0.109 | 0.97 (0.56, 1.67) | 0.913 | 1.37 (0.93, 2.02) | 0.112 | 0.54 (0.12, 2.56) | 0.443 |
| Global Biobank Meta-analysis Initiative | 1.79 (1.34, 2.41) | <0.001 | 1.51 (0.98, 2.32) | 0.060 | 1.79 (1.34, 2.41) | <0.001 | 10.29 (3.49, 30.36) | <0.001 |
| Meta-analysis (random effects) | 1.62 (1.25, 2.09) | <0.001 | 1.25 (0.81, 1.92) | 0.306 | 1.62 (1.25, 2.09) | <0.001 | 2.50 (0.14, 44.44) | 0.533 |

*ORs are expressed per one unit increase in log odds of smoking initiation.
†ORs are expressed per one SD increase (equivalent to an individual smoking 20 cigarettes a day for 15 years and stopping 17 years ago or an individual smoking 60 cigarettes a day for 13 years and stopping 22 years ago) in the lifetime smoking index.
IPF, idiopathic pulmonary fibrosis; MR, Mendelian randomization; PRESSO, pleiotropy residual sum and outlier.

of horizontal pleiotropy or heterogeneity for some associations (table 3). MR-PRESSO identified only rs6948707 at *MAD1L1* (a known IPF susceptibility signal) as an outlier and showed an effect directionally consistent with the IVW method after correction (table 2). Results were relatively stable in other sensitivity analyses although with wide confidence intervals. In the reverse MR analysis, genetic predisposition to IPF showed a null association with smoking initiation (OR=1.00; p=0.720) and lifetime smoking (beta=0.002; p=0.378), indicating the unidirectionality of the inferred relationship (online supplemental table 4). The leave-one-out analysis demonstrated that no single SNP (including rs6948707) substantially influenced the overall estimate (online supplemental figures 1–4). Scatter plots are presented in online supplemental figures 5–8.

## DISCUSSION

Because of the low incidence of IPF, to date only two cohort studies have evaluated the longitudinal association of smoking with IPF.[7 8] Both studies showed that smoking could increase the risk of IPF in a dose-response manner. However, a recent one-sample MR study reported that smoking is unlikely to be a causal factor for IPF, based on 871 IPF cases from the UK Biobank and an instrument constituted of 52 SNPs (explaining 0.7% of the variance in smoking volume).[9] The limited power of this study may have prevented detection of a causal effect. In contrast, our two-sample MR analysis had a much larger sample size (10382 IPF cases) and utilised stronger instruments that explain more variance in smoking, providing additional evidence for a potential causal effect of smoking on IPF. Several potential pathways may explain the role of cigarette smoke in the pathogenesis of IPF, including oxidative stress, inflammation, and telomere shortening.[10]

A common limitation in the MR setting is the presence of horizontal pleiotropy, which cannot be fully addressed even with sensitivity analyses based on different assumptions, because pleiotropy is widespread across the genome. Moreover, due to the use of summary data, we were unable to conduct a stratified analysis by smoking status or a nonlinear analysis to explore the threshold effect.

In conclusion, this study provides evidence for the potential causal effect of smoking on IPF. Further well-designed MR studies with more clinically diagnosed IPF cases are warranted to confirm our findings.

**Table 3** Heterogeneity, horizontal pleiotropy, and outlier tests

| Exposure | Outcome | IVW Q statistic (*P*-value) | MR-Egger intercept (*P*-value) | Outliers detected by MR-PRESSO |
|---|---|---|---|---|
| Smoking initiation | IPF (International IPF Genetics Consortium) | 516 (<0.001) | −0.005 (0.551) | rs6948707 |
| Smoking initiation | IPF (Global Biobank Meta-analysis Initiative) | 366 (<0.001) | −0.005 (0.436) | NA |
| Lifetime smoking | IPF (International IPF Genetics Consortium) | 140 (0.141) | 0.014 (0.229) | NA |
| Lifetime smoking | IPF (Global Biobank Meta-analysis Initiative) | 118 (0.262) | −0.034 (0.002) | NA |

IVW, inverse variance weighted; MR, Mendelian randomization; NA, not available; PRESSO, pleiotropy residual sum and outlier.

**Acknowledgements** The authors thank the International IPF Genetics Consortium and Global Biobank Meta-analysis Initiative for sharing the summary statistics for IPF.

**Contributors** JZ contributed to analysis and interpretation of data and drafting the work. DZ contributed to data interpretation. MY contributed to study design and analysis plan. YL contributed to acquisition and interpretation of data. All authors participated in revisions and approved the final version.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.



## REFERENCES

1 Martinez FJ, Collard HR, Pardo A, *et al*. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* 2017;3:17074.
2 Liu M, Jiang Y, Wedow R, *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;51:237–44.
3 Wootton RE, Richmond RC, Stuijfzand BG, *et al*. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med* 2020;50:2435–43.
4 Larsson SC, Burgess S. Appraising the causal role of smoking in multiple diseases: a systematic review and meta-analysis of Mendelian randomization studies. *EBioMedicine* 2022;82:104154.
5 Allen RJ, Stockwell A, Oldham JM, *et al*. Genome-wide association study across five cohorts identifies five novel Loci associated with idiopathic pulmonary fibrosis. *Thorax* 2022;77:829–33.
6 Zhou W, Kanai M, Wu K-HH, *et al*. Global Biobank meta-analysis initiative: Powering genetic discovery across human disease. *Cell Genom* 2022;2:100192.
7 Bellou V, Belbasis L, Evangelou E. Tobacco smoking and risk for pulmonary fibrosis: a prospective cohort study from the UK Biobank. *Chest* 2021;160:983–93.
8 Bae W, Lee C-H, Lee J, *et al*. Impact of smoking on the development of idiopathic pulmonary fibrosis: results from a nationwide population-based cohort study. *Thorax* 2022;77:470–6.
9 Duckworth A, Gibbons MA, Beaumont RN, *et al*. A Mendelian Randomisation study of smoking causality in IPF compared with COPD. *medRxiv* 2020.
10 Oh CK, Murray LA, Molfino NA. Smoking and idiopathic pulmonary fibrosis. *Pulm Med* 2012;2012:808260.