



## OPEN ACCESS

## EDITED BY

Jennie L. Williams,  
Stony Brook University, United States

## REVIEWED BY

Jelena Stojic,  
University of Belgrade, Serbia  
Meng Zhu,  
Nanjing Medical University, China

## \*CORRESPONDENCE

Lei Li

✉ lilei19830104@163.com

<sup>†</sup>These authors have contributed equally to this work and share first authorship

RECEIVED 12 January 2023

ACCEPTED 20 April 2023

PUBLISHED 03 May 2023

## CITATION

Wu H, Yang J, Wang H and Li L (2023) Mendelian randomization to explore the direct or mediating associations between socioeconomic status and lung cancer. *Front. Oncol.* 13:1143059. doi: 10.3389/fonc.2023.1143059

## COPYRIGHT

© 2023 Wu, Yang, Wang and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Mendelian randomization to explore the direct or mediating associations between socioeconomic status and lung cancer

Hong Wu<sup>1,2†</sup>, Jing Yang<sup>1†</sup>, Hui Wang<sup>3</sup> and Lei Li<sup>1\*</sup>

<sup>1</sup>Department of Oncology, Binzhou Medical University Hospital, Binzhou, Shandong, China,

<sup>2</sup>Department of Research, Guangxi Medical University Cancer Hospital, Nanning, Guangxi, China,

<sup>3</sup>Department of Orthopaedic Trauma and Hand Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China

**Objective:** The purpose of this study was to verify whether there are direct or mediated causal associations between socioeconomic status and lung cancer.

**Methods:** Pooled statistics were obtained from corresponding genome-wide association studies. The inverse-variance weighted, weighted median, MR-Egger, MR-PRESSO and contamination-mixture methods were used as supplements to Mendelian randomization (MR) statistical analysis. Cochran's Q value and the MR-Egger intercept were used for sensitivity analysis.

**Results:** In the univariate MR analysis, household income and education had protective effects on overall lung cancer (income:  $P = 5.46 \times 10^{-4}$ ; education:  $P = 4.79 \times 10^{-7}$ ) and squamous cell lung cancer (income:  $P = 2.67 \times 10^{-3}$ ; education:  $P = 1.42 \times 10^{-10}$ ). Smoking and BMI had adverse effects on overall lung cancer (smoking:  $P = 2.10 \times 10^{-7}$ ; BMI:  $P = 5.67 \times 10^{-4}$ ) and squamous cell lung cancer (smoking:  $P = 5.02 \times 10^{-6}$ ; BMI:  $P = 2.03 \times 10^{-7}$ ). Multivariate MR analysis found that smoking and education were independent risk factors for overall lung cancer (smoking:  $P = 1.96 \times 10^{-7}$ ; education:  $P = 3.11 \times 10^{-3}$ ), while smoking was an independent risk factor for squamous cell lung cancer ( $P = 2.35 \times 10^{-6}$ ). Smoking, education, and household income mediate the effect of BMI on overall lung cancer (smoking 50.0%, education 49.2%, income 25.3%) and squamous cell lung cancer (smoking 34.8%, education 30.8%, income 21.2%). Smoking, education, and BMI mediate the effect of income on overall lung cancer (smoking 13.9%, education 54.8%, BMI 9.4%) and squamous cell lung cancer (smoking 12.6%, education 63.3%, BMI 11.6%). Smoking, BMI, and income mediate the effect of education on squamous cell lung cancer (smoking 24.0%, BMI 6.2%, income 19.4%).

**Conclusion:** Income, education, BMI, and smoking are causally associated with both overall lung cancer and squamous cell lung cancer. Smoking and education are independent association factors for overall lung cancer, while smoking is an independent association factor for squamous cell lung cancer. Smoking and

education also play important mediating roles in overall lung cancer and squamous cell lung cancer. No causal relationship was found between multiple risk factors associated with socioeconomic status and lung adenocarcinoma.

#### KEYWORDS

lung cancer, socioeconomic, GWAS, SNP, Mendelian randomization

## Introduction

Cancer is the leading cause of death in countries around the world (1). According to Global Cancer Statistics 2020, lung cancer is the leading cause of cancer incidence and mortality in men, whereas it has the third highest incidence and the second highest mortality in women. Lung cancer, with a global incidence of 11.4%, ranks second among all new cancer cases. It remains the leading cause of cancer death (18%) (2). Socioeconomic status (SES) has been found to be associated with different types of cancer (3). It is a complex factor that can cover multiple dimensions of an individual's social and economic life circumstances and can be measured using information about education, income, and/or occupation (4). In addition, one study showed strong associations between SES, as measured by education level, household wealth and occupational rank, and smoking (5). A study of socioeconomic and alcohol consumption found that individuals with lower education levels and living in poverty (low income) were associated with higher levels of alcohol consumption (6). People in middle-income countries drink more alcohol than those in high-income countries (6). A cross-sectional study of a PERSIAN cohort of 20,000 Iranian adults found that the most important variables influencing higher body mass index (BMI) levels are SES (75.8%) and education level (-4.1%) (7). Another cross-sectional study based on data from the 2008 Canadian Community Health Survey found a slight association between occupational physical activity and BMI in women, while no association between occupational physical activity and BMI was detected in men (8). Therefore, SES is directly or indirectly associated with education, household income, smoking, alcohol consumption, BMI, and physical activity. This study will evaluate the causal relationship between SES and lung cancer from these dimensions.

The direct effect of smoking (both active and passive) on the development of lung cancer has been confirmed in many studies (9). Education level was an important independent predictor of lung cancer when appropriate follow-up was conducted for incidental findings of clinically significant lung cancer (10). A study based on population epidemiology in Japan found a strong positive association between alcohol consumption and lung cancer in women, but this association almost disappeared after adjusting for smoking (11). A lower BMI was also found to be associated with an increased risk of lung cancer in men in the Japanese population (12). However, physical activity was not found to

reduce the risk of lung cancer (13, 14). Earlier studies have shown an increased risk of lung cancer in low SES groups, and the association may be mediated by unexplained smoking exposure, lifestyle or occupational hazards (15, 16).

This study applied Mendelian randomization (MR) to explore factors potentially associated with SES, including education, household income, smoking, alcohol consumption, BMI, and physical activity, and to assess their association with lung cancer. MR is a data analysis method mainly used in epidemiological etiological inference in recent years. Traditional observational epidemiological studies have encountered many challenges in identifying disease etiology and inferring causality, such as reverse causal associations, potential confounders, and exposure factors with minor effects. MR, which uses genetic variation as instrumental variables (IVs) in the study of exposure factors to infer the association between gene-determined phenotypes and diseases, is not affected by the traditional observational interference factors mentioned above (17). Effective IV selection for MR requires the following three key assumptions. Relevance assumption is a genetic variation associated with risk factors for interest; in independence assumption, there are no unmeasurable confounding factors in the association between genetic variation and outcome; and in exclusion restriction, genetic variation cannot directly influence the outcome except through risk factors (18). In this study, univariate, multivariate and mediating MR methods were used to explore the association of factors related to SES with lung cancer. We further explored the direct or indirect role of risk factors and the proportion of indirect risk factors mediated by other factors.

## Material and methods

### Genome-wide association study summary statistics

Education level has certain heritability. James et al. conducted a large-scale genetic association analysis of educational attainment in a sample of approximately 1.1 million individuals and identified 1,271 independent genome-wide significant single nucleotide polymorphisms (SNPs). The study measured the number of years of schooling completed by the age of at least 30. All association analyses were performed in a sample limited to individuals of

European ancestry. The combination of education and the three related cognitive phenotypes explained 11–13% of the differences in educational attainment and 7–10% of the differences in cognitive performance (19).

Summary data on household income came from the consortium of the MRC-integrative epidemiology unit, which Ben Elsworth summarized in 2018. It included a 397,751 European population from the UK Biobank, with 9,851,869 SNPs linked to household income (<https://gwas.mrcieu.ac.uk/datasets/ukb-b-7408/>). The UK Biobank is a large prospective cohort study of approximately 500,000 adults (40–69 years of age) from 22 centers in the United Kingdom (20).

Liu et al. studied the genetic etiology of tobacco and alcohol use in up to 1.2 million Europeans and identified 566 genetic variants associated with multiple stages of tobacco use as well as alcohol use (21). Phenotypes associated with the initiation of smoking included the age at which regular smoking began and whether regular smoking occurred. Phenotypes associated with the severity of smoking were measured by the number of cigarettes smoked per day. The phenotype of smoking cessation was evaluated for previous smokers. Alcohol-related phenotypes were measured using weekly alcohol consumption (21). The World Health Organization defines a smoker as someone who has smoked continuously or cumulatively for six months or more during their lifetime. Therefore, the long-term behavior of regular smoking is more representative of an individual's smoking behavior than the age of initiation, the number of cigarettes smoked per day, and cessation. Our study included SNPs associated with regular smoking and weekly alcohol consumption as associated phenotypes for smoking and alcohol consumption, respectively.

Yengo et al. conducted a genome-wide association study (GWAS) of BMI in approximately 250,000 European participants and found approximately 100 independent SNPs. The results were pooled with a GWAS of BMI from the UK Biobank involving approximately 450,000 people of European ancestry. The meta-analysis reached 700,000 people and ultimately identified approximately 941 nearly independent BMI-related SNPs ( $P < 1 \times 10^{-8}$ ), which explained approximately 6% of the variance in BMI (22).

Klimentidis et al. identified multiple variants in a GWAS of habitual physical activity in more than 377,000 UK Biobank participants. The phenotypes associated with physical activity were divided into four categories, including moderate physical activity, vigorous physical activity, strenuous sports or other exercise, and accelerometer-based physical activity (23). Strenuous exercise or other exercise-related phenotypes were selected for physical activity exposure in this study. Because it was a review of the previous 4 weeks of physical activity, the assessed individuals spent 2–3 or more days per week in physical sports or other exercise, each time lasting 15–30 minutes or more.

The variants associated with lung cancer phenotypes in this study were obtained from the International Lung Cancer Consortium (ILCCO) using the MRBase database (<https://www.mrbase.org/>). Summary statistics were obtained from four GWAS meta-analyses of European ancestry: the MD Anderson Cancer Center (MDACC) GWAS, the Institute of Cancer Research

(ICR) GWAS, the National Cancer Institute (NCI) GWAS and the International Agency for Research on Cancer (IARC) GWAS (24). A total of 27,209 subjects were included (11,348 cases and 15,861 controls). Of the subtypes of lung cancer, 3,275 cases were defined as squamous cell lung cancer, and 3,442 cases were lung adenocarcinoma. There was no sample overlap between the exposures and outcomes selected for this study.

## Mendelian randomization statistical analysis

In this study, univariate MR analysis was used to verify the causal relationship between multiple factors related to SES and lung cancer. Multivariate MR was used to distinguish whether risk factors were independent after evidence showed that multiple risk factors were associated with lung cancer. The independent risk factor was further analyzed by mediating Mendelian randomization to explore the proportion of it mediated by other risk factors. The proportion of the effect that is mediated by any of the potential mediators was estimated using the following equation (standard error estimated using the error propagation method):

$$E(\%) = \frac{\sum_{k=1}^K \beta_1 \times \beta_2_k}{\sum_{k=1}^K \beta_3 + \beta_1 \times \beta_2_k}$$

where the regression coefficient  $\beta_1$  is the MR effect of the independent risk factor on the mediator,  $\beta_2$  is the MR effect of the mediator with lung cancer adjusted for the independent risk factor, and  $\beta_3$  is the MR effect of the independent risk factor on lung cancer adjusted for the potential mediator. All regression coefficients were derived from MR instrumental analysis using the inverse-variance weighted (IVW) method, assuming no correlation between the mediators (25–27).

Genome-wide significant single nucleotide polymorphisms (SNPs) ( $P < 5 \times 10^{-8}$ ) associated with SES risk factors (education, household income, smoking, alcohol consumption, BMI, and physical activity) were selected as IVs. SNPs selected as IVs with linkage disequilibrium ( $r^2 = 0.001$  and  $KB = 10000$ ) and a minor allele frequency of less than 0.01 were removed. Phenotypic variation explained by SNPs was calculated using the formula:  $R^2 = 2 \times \text{beta}^2 \times (1 - \text{EAF}) \times \text{EAF} / \text{SD}^2$  (EAF, effect allele frequency; SD, standard deviation; beta refers to the effect of each SNP on the exposure) (28). The weak instrumental variable bias of SNPs was evaluated by the F statistic  $((N - k - 1) / k) \times (R^2 / (1 - R^2))$ ; N, the sample size; k, number of SNPs) (29). SNPs with an association strength greater than 10 with exposure phenotypes were included in the study.

IVW was used as the primary analytical method to evaluate the effect between exposure and outcome (30). The weighted median (31), MR-Egger (32), MR-PRESSO (33) and contamination-mixture (34) methods were used as supplements to MR statistical analysis. Cochran's Q value (35) and the MR-Egger intercept (32) were used for sensitivity analysis to assess heterogeneity and horizontal pleiotropy. When horizontal pleiotropy existed, MR-PRESSO could remove outliers based on the IVW method to provide an estimation of the causal effect again (33). When neither heterogeneity nor horizontal pleiotropy existed, the IVW method was considered the primary assessment method. When the

weighted median and IVW were in the same direction and there was heterogeneity, the results of the weighted median method were accepted. Because the weighted median estimate provides an unbiased effect even if up to 50% of genetic IVs are invalid, IVW requires that all SNPs used as IVs are valid (31). To conclude, for there to be a causal relationship between exposure and outcome, the P value of the results of MR analysis should be less than the significance level of  $8.33 \times 10^{-3}$  corrected by Bonferroni (P value threshold = 0.05/6, corrected for 6 pairs of exposure and outcome).

All MR analyses were performed using the “TwoSampleMR (36)”, “MR-PRESSO (33)” and “MendelianRandomization (37)” packages of R software version 4.1.1.

## Results

The number of SNPs as instrumental variables for the six risk factors associated with SES ranged from 14 to 507. The total explained variance and F statistics of the selected instrumental variables are shown in [Supplementary Table S1](#). Genome-wide significant SNPs for the six risk factors associated with SES that were extracted as IVs are characterized in [Supplemental Tables S2–7](#).

In the univariate MR analysis, lung cancer as the outcome was divided into three categories: lung cancer (including adenocarcinoma and squamous cell carcinoma), lung adenocarcinoma, and squamous cell lung cancer. There was evidence that household income and education had protective effects on overall lung cancer in the main IVW method (income: OR 0.55, 95% CI 0.39–0.77,  $P = 5.46 \times 10^{-4}$ ; education: OR 0.64, 95% CI 0.53–0.76,  $P = 4.79 \times 10^{-7}$ ). Smoking and BMI had adverse effects on overall lung cancer (smoking: OR 1.58, 95% CI 1.33–1.88,  $P = 2.10 \times 10^{-7}$ ; BMI: OR 1.23, 95% CI 1.09–1.38,  $P = 5.67 \times 10^{-4}$ ). There was no significant association between alcohol consumption, strenuous sports and overall lung cancer (alcohol drink: OR 1.52, 95% CI 1.11–2.08,  $P = 9.45 \times 10^{-3}$ ; strenuous sports: OR 0.21, 95% CI

0.04–1.04,  $P = 5.56 \times 10^{-2}$ ;  $P$  value <  $8.33 \times 10^{-3}$  corrected by Bonferroni) ([Figure 1](#)). Income, education, smoking, alcohol consumption, BMI and strenuous sports were not significantly associated with lung adenocarcinoma (income: OR 0.54, 95% CI 0.33–0.89,  $P = 1.57 \times 10^{-2}$ ; education: OR 0.76, 95% CI 0.58–0.99,  $P = 3.95 \times 10^{-2}$ ; smoking: OR 1.38, 95% CI 1.06–1.80,  $P = 1.80 \times 10^{-2}$ ; alcohol drink: OR 1.31, 95% CI 0.65–2.65,  $P = 4.56 \times 10^{-1}$ ; BMI: OR 0.93, 95% CI 0.78–1.11,  $P = 4.27 \times 10^{-1}$ ; strenuous sports: OR 0.21, 95% CI 0.02–2.30,  $P = 2.02 \times 10^{-1}$ ;  $P$  value <  $8.33 \times 10^{-3}$  corrected by Bonferroni) ([Figure 1](#)). There was evidence that household income and education had protective effects against squamous cell lung cancer (income: OR 0.47, 95% CI 0.29–0.77,  $P = 2.67 \times 10^{-3}$ ; education: OR 0.42, 95% CI 0.32–0.55,  $P = 1.42 \times 10^{-10}$ ). Smoking and BMI had adverse effects on squamous cell lung cancer (smoking: OR 1.84, 95% CI 1.42–2.39,  $P = 5.02 \times 10^{-6}$ ; BMI: OR 1.58, 95% CI 1.33–1.87,  $P = 2.03 \times 10^{-7}$ ). There was no significant association between alcohol consumption, strenuous sports and squamous cell lung cancer (alcohol drink: OR 1.58, 95% CI 0.98–2.56,  $P = 5.94 \times 10^{-2}$ ; strenuous sports: OR 0.04, 95% CI 0.002–0.83,  $P = 3.69 \times 10^{-2}$ ;  $P$  value <  $8.33 \times 10^{-3}$  corrected by Bonferroni) ([Figure 1](#)). Heterogeneity and horizontal pleiotropy were not observed in the Mendelian randomized sensitivity analysis.

The results of the IVW, weighted median, MR–Egger, MR–PRESSO and contamination-mixture methods for the three types of lung cancer, as well as the sensitivity analysis results of Cochran’s Q and the MR–Egger intercept, are shown in [Supplemental Tables S8–10](#).

Four risk factors - BMI, smoking, education, and income - had causal relationships with overall lung cancer and squamous cell lung cancer. Multivariate MR analysis found that smoking and education were independent risk factors for overall lung cancer (smoking: OR 1.79, 95% CI 1.44–2.23,  $P = 1.96 \times 10^{-7}$ ; education: OR 0.53, 95% CI 0.34–0.81,  $P = 3.11 \times 10^{-3}$ ), while smoking was an independent risk factor for squamous cell lung cancer (OR 2.17, 95% CI 1.57–3.00,  $P = 2.35 \times 10^{-6}$ ) ([Figure 2](#)). BMI and income were nonindependent risk factors for overall lung cancer, while BMI,

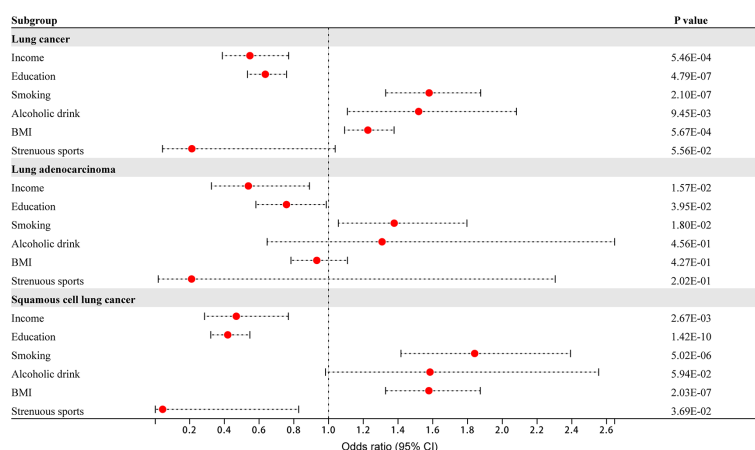


FIGURE 1 The effect of 6 dimensional factors of socioeconomic status on the risk of lung cancer and its subtypes.

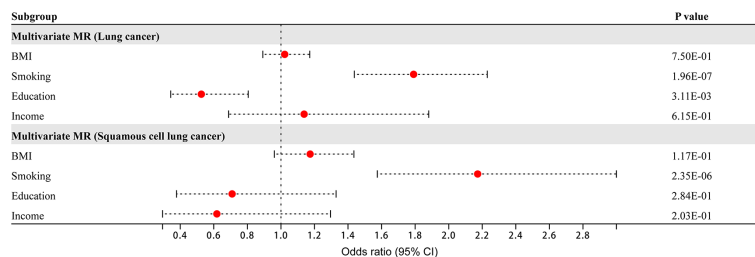


FIGURE 2 Multivariate Mendelian randomization analysis of overall lung cancer and squamous cell lung cancer.

education and income were nonindependent risk factors for squamous cell lung cancer. These nonindependent risk factors were mediated by other risk factors.

Mediating MR was used to assess the mediating effect of nonindependent risk factors on overall lung cancer through other risk factors, the results of which are shown in Figure 3. The effect of BMI on overall lung cancer risk was attenuated from an OR of 1.23 (95% CI 1.09–1.39) to an OR of 1.12 (95% CI 0.98–1.27) after adjusting for smoking, to an OR of 1.12 (95% CI 0.98–1.28) after adjusting for education, to an OR of 1.18 (95% CI 1.04–1.35) after adjusting for income, and to an OR of 1.02 (95% CI 0.89–1.17) after adjusting for all three factors. The mediating percentages of different risk factors between BMI and overall lung cancer risk were smoking at 50.0% (95% CI 30.7–69.3%), education at 49.2% (95% CI 30.3–68.1%), income at 25.3% (95% CI 10.6–40.0%), and all three factors at 41.3% (95% CI 31.3–51.3%). The effect of income on overall lung cancer risk was attenuated from an OR of 0.44 (95% CI 0.30–0.66) to an OR of 0.49 (95% CI 0.33–0.74) after adjusting for smoking, to an OR of 0.77 (95% CI 0.46–1.31) after adjusting for education, to an OR of 0.66 (95% CI 0.45–0.97) after adjusting for BMI, and to an OR of 1.14 (95% CI 0.69–1.88) after adjusting for all three factors. The mediating percentages of different risk factors between household income and overall lung cancer risk were as follows: smoking at 13.9% (95% CI 7.8–20.0%), education at 54.8% (95% CI 16.4–93.2%), BMI at 9.4% (95% CI 3.2–15.6%), and all three factors at 25.5% (95% CI 15.6–35.4%).

The mediating effect of nonindependent risk factors on squamous cell lung cancer through other risk factors is shown in Figure 4. The effect of BMI on squamous cell lung cancer risk was attenuated from an OR of 1.54 (95% CI 1.30–1.83) to an OR of 1.32 (95% CI 1.10–1.58) after adjusting for smoking, to an OR of 1.34 (95% CI 1.10–1.62) after adjusting for education, to an OR of 1.41 (95% CI 1.17–1.70) after adjusting for income, and to an OR of 1.17 (95% CI 0.96–1.44) after adjusting for all three factors. The mediating percentages of different risk factors between BMI and squamous cell lung cancer risk were as follows: smoking at 34.8% (95% CI 23.3–46.3%), education at 30.8% (95% CI 20.4–41.2%), income at 21.2% (95% CI 10.9–31.5%), and all three factors at 28.9% (95% CI 22.7–35.1%). The effect of education on squamous cell lung cancer risk was attenuated from an OR of 0.40 (95% CI 0.30–0.52) to an OR of 0.50 (95% CI 0.37–0.68) after adjusting for smoking, to an OR of 0.39 (95% CI 0.27–0.57) after adjusting for BMI, to an OR of 0.47 (95% CI 0.26–0.84) after adjusting for income, and to an OR of 0.71 (95% CI 0.38–1.33) after adjusting for all three factors. The mediating percentages of different risk factors between education and squamous cell lung cancer risk were as follows: smoking at 24.0% (95% CI 15.3–32.7%), BMI at 6.2% (95% CI 3.5–8.9%), income at 19.4% (95% CI -9.9–48.7%), and all three factors at 16.3% (95% CI 6.3–26.3%). The effect of income on squamous cell lung cancer risk was attenuated from an OR of 0.41 (95% CI 0.25–0.69) to an OR of 0.42 (95% CI 0.23–0.76) after adjusting for smoking, to an OR of 0.74 (95% CI 0.33–1.65) after adjusting

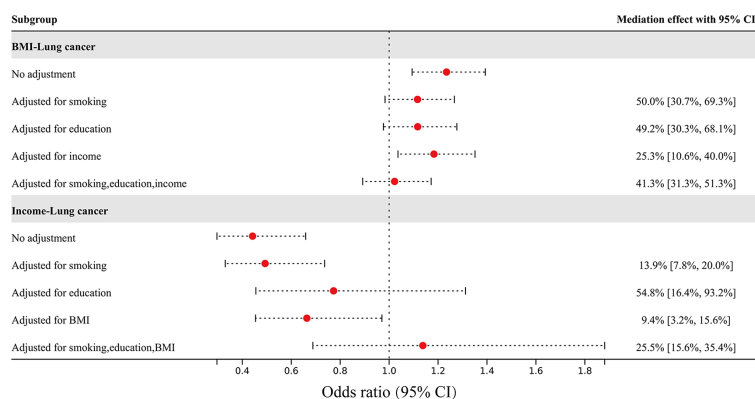


FIGURE 3 The effect of BMI and income on overall lung cancer risk after adjusting for mediators, and the percentage of mediating factors.



for education, to an OR of 0.52 (95% CI 0.30–0.88) after adjusting for BMI, and to an OR of 0.62 (95% CI 0.30–1.30) after adjusting for all three factors. The mediating percentages of different risk factors between income and squamous cell lung cancer risk were smoking at 12.6% (95% CI 6.0–19.2%), education at 63.3% (95% CI 19.4–100%), BMI at 11.6% (95% CI 5.2–18.0%), and all three factors at 28.7% (95% CI 17.6–39.8%). The data used in the mediated MR analysis for multiple independent risk factors are shown in Supplemental Tables S11–13.

## Discussion

In this study, we found that household income, education, smoking, and BMI related to SES were associated with both overall lung cancer and squamous cell lung cancer. There was no evidence of a causal relationship between the factors selected as SES and lung adenocarcinoma. Smoking and education were independently associated with overall lung cancer, while smoking was an independent risk factor for squamous cell lung cancer.

Consistent with our findings, education has been found to be associated with lung cancer in multiple studies. The 2021 National Health Survey of the United States found that lung cancer patients aged 45 or older had the highest number of education degrees below high school (8.6%). The rate of lung cancer decreased with increasing education level, with the lowest prevalence among those with a bachelor’s degree or higher (1.6%) (38). Another cohort study of Norwegians also found that low education increased the risk of dying from lung cancer (39). A Mendelian randomized study on education and lung cancer also showed that low education was a causal risk factor in the development of lung cancer (40). There is much research evidence on the causal relationship between smoking and lung cancer. Smoking is an important risk factor for lung cancer and is independently associated with a higher risk of lung cancer (41, 42). Smoking cessation can improve lung cancer incidence and survival rates, and

even those who quit in middle age can avoid most of the risk of lung cancer (43–46). An observational trial of the mediating role of smoking in the relationship between education and lung cancer found that the indirect effects of smoking varied by level of education, with the strongest effects observed for those with the least education (47). Therefore, in this study, we also demonstrated causal relationships between smoking, education and overall lung cancer. However, neither smoking nor education was an independent risk factor for squamous cell lung cancer subtypes.

Associations of BMI with lung cancer have been reported across ethnic groups and across observational cohort studies; however, their findings are ambiguous. Based on relevant studies of the International Lung Cancer Consortium and the Chinese population, it was found that both an increase and a decrease in BMI could increase the risk of lung cancer and reduce the survival rate of lung cancer (48, 49). Decreased BMI was also found to be associated with poorer overall survival of lung cancer, while weight gain increased the risk of death but was not statistically significant (50). Other studies have found a positive correlation between BMI and the incidence of lung cancer or lung adenocarcinoma (51, 52). Additionally, an observational study showed that higher BMI was associated with a lower risk of lung cancer (53). From this, we can see that the results of observational studies vary and that racial differences may be one reason for this difference (54, 55). We included people of European descent in this study and concluded that an increase in BMI increases the risk of both overall lung cancer and squamous cell lung cancer. However, the causal relationship between BMI and lung cancer was not an independent risk factor; it was mediated by smoking, education and income, with smoking having the strongest mediating effect, followed by education. Another MR study also found a mediating role of smoking in the association between BMI and lung cancer (56).

Overall survival of lung cancer was positively correlated with high income of patients and was not correlated with race (57). A lower economic gradient was associated with higher mortality from lung cancer (58). Similarly, studies found that the prevalence and mortality

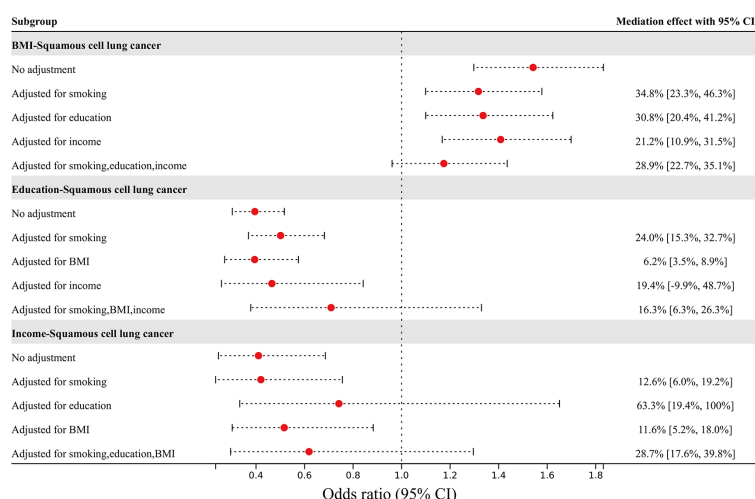


FIGURE 4 The effects of BMI, education, and income on squamous cell lung cancer risk after adjusting for mediators, and the percentage of mediating factors.

of lung cancer were higher among people with economic poverty and low education levels (59, 60). In this study, household income was negatively correlated with overall lung cancer and squamous cell lung cancer, but it was not an independent association factor. Education, smoking and BMI play a mediating role between the two, with education playing the strongest mediating role, followed by smoking. The higher the level of education, the lower the overall risk of lung cancer. Education was an independent association factor for overall lung cancer but not squamous cell lung cancer. The negative correlation between education and squamous cell lung cancer was mediated by smoking, income and BMI, among which smoking was the most important mediator.

Mendelian randomization has some limitations that need to be noted. First, exposures and outcomes can only be selected based on the limited GWAS data available, which has certain limitations. In addition, estimates of intermediary proportions may be biased due to the noncollapsibility of ORs (27). Finally, the study was based only on populations of European descent. This conclusion only applies to European ancestry and cannot represent all ethnic groups. Further verification is needed for other ancestry groups.

## Conclusion

Income, education, BMI, and smoking are causally associated with both overall lung cancer and squamous cell lung cancer. Smoking and education are independent association factors for overall lung cancer, while smoking is an independent association factor for squamous cell lung cancer. Second, smoking and education also play important mediating roles in lung cancer: the causal association between BMI and overall lung cancer and squamous cell lung cancer is mainly mediated by smoking and education. The causal association between household income and overall lung cancer and squamous cell lung cancer is mainly mediated by education; finally, the causal association between education and squamous cell lung cancer is mainly mediated by smoking. No causal relationship is found between multiple risk factors associated with socioeconomic status and lung adenocarcinoma.

## References

1. Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer* (2021) 127(16):3029–30. doi: 10.1002/cncr.33587
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
3. Woods LM, Rachet B, Coleman MP. Origins of socio-economic inequalities in cancer survival: a review. *Ann Oncol* (2006) 17(1):5–19. doi: 10.1093/annonc/mdj007
4. Quaglia A, Lillini R, Mamo C, Ivaldi E, Vercelli M. Socio-economic inequalities: a review of methodological issues and the relationships with cancer survival. *Crit Rev Oncol Hematol* (2013) 85(3):266–77. doi: 10.1016/j.critrevonc.2012.08.007
5. Schaap MM, Kunst AE. Monitoring of socio-economic inequalities in smoking: learning from the experiences of recent scientific studies. *Public Health* (2009) 123(2):103–9. doi: 10.1016/j.puhe.2008.10.015
6. Huckle T, Romeo JS, Wall M, Callinan S, Holmes J, Meier P, et al. Socio-economic disadvantage is associated with heavier drinking in high but not middle-

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Author contributions

HWu and JY drafted the manuscript and contributed to the conception of the study. HWa helped perform the analysis with constructive discussions. LL gave critical feedback and approved the final version. All authors contributed to the article and approved the submitted version.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1143059/full#supplementary-material>

income countries participating in the international alcohol control study. *Drug Alcohol Rev* (2018) 37 Suppl 2:S63–71. doi: 10.1111/dar.12810

7. Pourfarzi F, Rezaei S, Zahirian Moghadam T, Zandian H, Dibazar F. The socio-economic inequality in body mass index: a PERSIAN cohort-based cross-sectional study on 20,000 Iranian adults. *BMC Endocr Disord* (2022) 22(1):178. doi: 10.1186/s12902-022-01096-2

8. Barberio A, McLaren L. Occupational physical activity and body mass index (BMI) among Canadian adults: does physical activity at work help to explain the socio-economic patterning of body weight? *Can J Public Health* (2011) 102(3):169–73. doi: 10.1007/BF03404888

9. Shankar A, Dubey A, Saini D, Singh M, Prasad CP, Roy S, et al. Environmental and occupational determinants of lung cancer. *Transl Lung Cancer Res* (2019) 8(Suppl 1):S31–49. doi: 10.21037/tlcr.2019.03.05

10. Kapoor S, Deppen SA, Paulson AB, Haddad D, Cook JP, Sandler KL. Education level predicts appropriate follow-up of incidental findings from lung cancer screening. *J Am Coll Radiol* (2020) 17(5):613–9. doi: 10.1016/j.jacr.2019.12.014

11. Wakai K, Nagata C, Mizoue T, Tanaka K, Nishino Y, Tsuji I, et al. Alcohol drinking and lung cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Jpn J Clin Oncol* (2007) 37(3):168–74. doi: 10.1093/jcco/hyl1146
12. Abe SK, Narita S, Saito E, Sawada N, Shimazu T, Goto A, et al. Body mass index, height, weight change, and subsequent lung cancer risk: the Japan public health center-based prospective study. *Cancer Epidemiol Biomarkers Prev* (2021) 30(9):1708–16. doi: 10.1158/1055-9965.EPI-21-0195
13. Baumeister SE, Leitzmann MF, Bahls M, Meisinger C, Amos CI, Hung RJ, et al. Physical activity does not lower the risk of lung cancer. *Cancer Res* (2020) 80(17):3765–9. doi: 10.1158/0008-5472.CAN-20-1127
14. Xian W, Shen J, Zhou H, Liu J, Zhang Y, Zhang Z, et al. Mendelian randomization study indicates lack of causal relationship between physical activity and lung cancer. *J Cancer Res Clin Oncol* (2021) 147(1):177–81. doi: 10.1007/s00432-020-03409-1
15. Denton EJ, Hart D, Russell PA, Wright G, Conron M. Lung cancer and socioeconomic status: inextricably linked to place of residence. *Intern Med J* (2017) 47(5):563–9. doi: 10.1111/imj.13376
16. Ekberg-Aronsson M, Nilsson PM, Nilsson JA, Pehrsson K, Löfdahl CG. Socio-economic status and lung cancer risk including histologic subtyping—a longitudinal study. *Lung Cancer* (2006) 51(1):21–9. doi: 10.1016/j.lungcan.2005.08.014
17. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* (2003) 32(1):1–22. doi: 10.1093/ije/dyg070
18. Davies NM, Holmes MV, Davey Smith G. Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* (2018) 362:k601. doi: 10.1136/bmj.k601
19. Lee JJ, Wedow R, Okbay A, Kong E, Maghziyan O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* (2018) 50(8):1112–21. doi: 10.1038/s41588-018-0147-3
20. 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(3):e1001779. doi: 10.1371/journal.pmed.1001779
21. 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. doi: 10.1038/s41588-018-0307-5
22. Yengo L, Sidorenko J, Kempner KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Hum Mol Genet* (2018) 27(20):3641–9. doi: 10.1093/hmg/ddy271
23. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, et al. Genome-wide association study of habitual physical activity in over 377,000 UK biobank participants identifies multiple variants including *CADM2* and *APOE*. *Int J Obes (Lond)* (2018) 42(6):1161–76. doi: 10.1038/s41366-018-0120-3
24. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in *BRCA2* and *CHEK2* affect risk of lung cancer. *Nat Genet* (2014) 46(7):736–41. doi: 10.1038/ng.3002
25. Varbo A, Benn M, Smith GD, Timpson NJ, Tybjaerg-Hansen A, Nordestgaard BG. Remnant cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease. *Circ Res* (2015) 116(4):665–73. doi: 10.1161/CIRCRESAHA.116.304846
26. Xu L, Borges MC, Hemani G, Lawlor DA. The role of glycaemic and lipid risk factors in mediating the effect of BMI on coronary heart disease: a two-step, two-sample mendelian randomisation study. *Diabetologia* (2017) 60(11):2210–20. doi: 10.1007/s00125-017-4396-y
27. Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *Eur J Epidemiol* (2021) 36(5):465–78. doi: 10.1007/s10654-021-00757-1
28. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 caucasians. *PLoS One* (2015) 10(4):e0120758. doi: 10.1371/journal.pone.0120758
29. Burgess S, Thompson SG. Bias in causal estimates from mendelian randomization studies with weak instruments. *Stat Med* (2011) 30(11):1312–23. doi: 10.1002/sim.4197
30. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* (2013) 37(7):658–65. doi: 10.1002/gepi.21758
31. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* (2016) 40(4):304–14. doi: 10.1002/gepi.21965
32. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *Int J Epidemiol* (2015) 44(2):512–25. doi: 10.1093/ije/dyv080
33. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* (2018) 50(5):693–8. doi: 10.1038/s41588-018-0099-7
34. Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust and efficient method for mendelian randomization with hundreds of genetic variants. *Nat Commun* (2020) 11(1):376. doi: 10.1038/s41467-019-14156-4
35. 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. doi: 10.1002/sim.6835
36. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife* (2018) 7:e34408. doi: 10.7554/eLife.34408
37. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing mendelian randomization analyses using summarized data. *Int J Epidemiol* (2017) 46(6):1734–9. doi: 10.1093/ije/dyx034
38. Julie D, Weeks PhD, Nazik Elgaddal MS. QuickStats: percentage of adults aged ≥45 years who have ever had lung cancer, by education level - national health interview survey, united states, 2021. *MMWR Morb Mortal Wkly Rep* (2022) 71(45):1460. doi: 10.15585/mmwr.mm7145a6
39. Hansen MS, Licaj I, Braaten T, Langhammer A, Le Marchand L, Gram IT. Smoking related lung cancer mortality by education and sex in Norway. *BMC Cancer* (2019) 19(1):1132. doi: 10.1186/s12885-019-6330-9
40. Zhou H, Zhang Y, Liu J, Yang Y, Fang W, Hong S, et al. Education and lung cancer: a mendelian randomization study. *Int J Epidemiol* (2019) 48(3):743–50. doi: 10.1093/ije/dyz121
41. Zhang P, Chen PL, Li ZH, Zhang A, Zhang XR, Zhang YJ, et al. Association of smoking and polygenic risk with the incidence of lung cancer: a prospective cohort study. *Br J Cancer* (2022) 126(11):1637–46. doi: 10.1038/s41416-022-01736-3
42. Romaszko-Wojtowicz A, Lorenc A, Buciński A, Doboszyńska A. Effects of tobacco smoking on the survivability of patients with multiple cancers and single lung cancer. *Int J Environ Res Public Health* (2022) 19(15):9179. doi: 10.3390/ijerph19159179
43. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *BMJ* (2000) 321(7257):323–9. doi: 10.1136/bmj.321.7257.323
44. Heiden BT, Engelhardt KE, Cao C, Meyers BF, Puri V, Cao Y, et al. Association between lung cancer screening and smoking cessation. *Cancer Epidemiol* (2022) 79:102194. doi: 10.1016/j.canep.2022.102194
45. Su Z, Jia XH, Zhao FH, Zhou QH, Fan YG, Qiao YL. Effect of time since smoking cessation on lung cancer incidence: an occupational cohort with 27 follow-up years. *Front Oncol* (2022) 12:817045. doi: 10.3389/fonc.2022.817045
46. Caini S, Del Riccio M, Vettori V, Scotti V, Martinoli C, Raimondi S, et al. Quitting smoking At or around diagnosis improves the overall survival of lung cancer patients: a systematic review and meta-analysis. *J Thorac Oncol* (2022) 17(5):623–36. doi: 10.1016/j.jtho.2021.12.005
47. Menvielle G, Franck JE, Radoi L, Sanchez M, Févotte J, Guizard AV, et al. Quantifying the mediating effects of smoking and occupational exposures in the relation between education and lung cancer: the ICARE study. *Eur J Epidemiol* (2016) 31(12):1213–21. doi: 10.1007/s10654-016-0182-2
48. Shepshelovich D, Xu W, Lu L, Fares A, Yang P, Christiani D, et al. Body mass index (BMI), BMI change, and overall survival in patients with SCLC and NSCLC: a pooled analysis of the international lung cancer consortium. *J Thorac Oncol* (2019) 14(9):1594–607. doi: 10.1016/j.jtho.2019.05.031
49. Wu Z, Xie S, Wang F, Chen S, Su K, Li F, et al. BMI changes and the risk of lung cancer in male never-smokers: a prospective cohort study. *Cancer Med* (2022) 11(5):1336–46. doi: 10.1002/cam4.4546
50. Yuan Q, Du M, Loehrer E, Johnson BE, Gainor JF, Lanuti M, et al. Postdiagnosis BMI change is associated with non-small cell lung cancer survival. *Cancer Epidemiol Biomarkers Prev* (2022) 31(1):262–8. doi: 10.1158/1055-9965.EPI-21-0503
51. Wood AM, Jonsson H, Nagel G, Häggström C, Manjer J, Ulmer H, et al. The inverse association of body mass index with lung cancer: exploring residual confounding, metabolic aberrations and within-person variability in smoking. *Cancer Epidemiol Biomarkers Prev* (2021) 30(8):1489–97. doi: 10.1158/1055-9965.EPI-21-0058
52. Jiang L, Sun YQ, Brumpton BM, Langhammer A, Chen Y, Mai XM. Body mass index and incidence of lung cancer in the HUNT study: using observational and mendelian randomization approaches. *BMC Cancer* (2022) 22(1):1152. doi: 10.1186/s12885-022-10215-0
53. Zhu H, Zhang S. Body mass index and lung cancer risk in never smokers: a meta-analysis. *BMC Cancer* (2018) 18(1):635. doi: 10.1186/s12885-018-4543-y
54. Zhao J, Barta JA, McIntire R, Shusted C, Zeigler-Johnson C, Juon HS. Racial difference in BMI and lung cancer diagnosis: analysis of the national lung screening trial. *BMC Cancer* (2022) 22(1):797. doi: 10.1186/s12885-022-09888-4
55. Braillon A. Lung cancer outcomes: are BMI and race clinically relevant? *Lung Cancer* (2021) 154:224. doi: 10.1016/j.lungcan.2020.12.021
56. Zhou W, Liu G, Hung RJ, Haycock PC, Aldrich MC, Andrew AS, et al. Causal relationships between body mass index, smoking and lung cancer: univariable and multivariable mendelian randomization. *Int J Cancer* (2021) 148(5):1077–86. doi: 10.1002/ijc.33292



57. Varlotto JM, Voland R, McKie K, Flickinger JC, DeCamp MM, Maddox D, et al. Population-based differences in the outcome and presentation of lung cancer patients based upon racial, histologic, and economic factors in all lung patients and those with metastatic disease. *Cancer Med* (2018) 7(4):1211–20. doi: 10.1002/cam4.1430
58. Polak M, Genowska A, Szafraniec K, Fryc J, Jamiolkowski J, Pająk A. Area-based socio-economic inequalities in mortality from lung cancer and respiratory diseases. *Int J Environ Res Public Health* (2019) 16(10):1791. doi: 10.3390/ijerph16101791
59. Hajizadeh M, Johnston GM, Manos D. Socio-economic inequalities in lung cancer incidence in Canada, 1992-2010: results from the Canadian cancer registry. *Public Health* (2020) 185:189–95. doi: 10.1016/j.puhe.2020.04.023
60. Redondo-Sánchez D, Petrova D, Rodríguez-Barranco M, Fernández-Navarro P, Jiménez-Moleón JJ, Sánchez MJ. Socio-economic inequalities in lung cancer outcomes: an overview of systematic reviews. *Cancers (Basel)* (2022) 14(2):398. doi: 10.3390/cancers14020398