Observational Study

Medicine

OPEN

Causal validation of the relationship between air pollution and lung cancer

A bidirectional Mendelian randomization study and meta-analysis

Xiaomin Wang, MM[a], Guihua Xiao, MM[b], Wanxian Xu, MM[a], Changguo Ni, MM[a,*]

Abstract

Recent studies suggest a link between air pollution and lung cancer, but causality remains uncertain due to confounding and reverse causation. Mendelian randomization (MR) reduces such bias and offers a new way to explore this relationship. MR is a method that uses genetic variants as instrumental variables to assess the causal relationship between an exposure and an outcome, effectively controlling for confounding and reverse causation. The inverse-variance weighted method is a commonly used approach in MR analysis, which estimates the overall causal effect by weighting the effect ratios of multiple single nucleotide polymorphisms, assuming all instruments are valid. Based on 2-sample MR, this study incorporated 5 air pollution indices and conducted MR analyses with lung cancer outcome data from 2 different sources. Subsequently, a meta-analysis was performed on the primary inverse-variance weighted results, followed by multiple corrections of the thresholds after the meta-analysis to ensure accuracy. Finally, reverse causality was tested through MR analysis for air pollution indices significantly associated with lung cancer. And the selection criteria for instrumental variables were: $P < 5 \times 10^{-6}$, $F > 10$, minor allele frequency > 0.01, clump_kb = 10,000, and clump_r^2 = 0.001. Five air pollution indices were analyzed using MR analysis and meta-analysis with lung cancer data from the FinnGen R12 and OpenGWAS databases. Multiple corrections were applied to the significance threshold results after the meta-analysis. The final results showed that only nitrogen dioxide (NO_2) exhibited a significant association, with an OR of 3.426 (95% CI: 1.897–6.186, $P = 2.21 \times 10^{-4}$). Additionally, the positive air pollution index NO_2 showed no evidence of reverse causality with lung cancer from either data source. This study demonstrates a significant causal association between NO_2 and lung cancer, indicating that NO_2 may be a potential risk factor for lung cancer.

Abbreviations: CI = confidence interval, eaf = effect allele frequency, GWAS = genome-wide association study, IVs = instrumental variables, IVW = inverse-variance weighted, Kb = kilobase pairs, LD = linkage disequilibrium, MAF = minor allele frequency, MR = Mendelian randomization, NO_2 = nitrogen dioxide, NSAIDs = nonsteroidal anti-inflammatory drugs, OR = odds ratio, SNPs = single nucleotide polymorphisms.

Keywords: air pollution 1, lung cancer 2, mendelian randomization analysis 3, reverse mendelian randomization analysis 4

Health Research Project of the Kunming Municipal Health Commission, Project Number: 2022-04-02-006.

Participants in FinnGen provided informed consent for biobank research on basis of the Finnish Biobank Act. Alternatively, separate research cohorts, collected before the Finnish Biobank Act came into effect (in September 2013) and the start of FinnGen (August 2017) were collected on the basis of study-specific consent and later transferred to the Finnish biobanks after approval by Fimea, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol (number HUS/990/2017).

The authors have no conflicts of interest to disclose.

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

The FinnGen study is approved by the THL (approval number THL/2031/6.02.00/2017, amendments THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019 and THL/1721/5.05.00/2019), the Digital and Population Data Service Agency (VRK43431/2017-3, VRK/6909/2018-3, and VRK/4415/2019-3), the Social Insurance Institution (KELA) (KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, and KELA 98/522/2019), and Statistics Finland (TK-53-1041-17).

Supplemental Digital Content is available for this article.

[a] The First People's Hospital of Kunming City and Calmette Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China, [b] Zhoupu Hospital, Pudong New District, Shanghai, China.

** Correspondence: Changguo Ni, The First People's Hospital of Kunming City and Calmette Affiliated Hospital of Kunming Medical University,1228 Beijing Road, Panglong District, Kunming 650032, Yunnan, China (e-mail: chngooni@163.com).*

How to cite this article: Wang X, Xiao G, Xu W, Ni C. Causal validation of the relationship between air pollution and lung cancer: A bidirectional Mendelian randomization study and meta-analysis. Medicine 2025;104:21(e42450).

Received: 20 January 2025 / Received in final form: 22 April 2025 / Accepted: 25 April 2025

http://dx.doi.org/10.1097/MD.0000000000042450

1. Introduction

Lung cancer is one of the most common and deadliest malignancies worldwide. In 2020, approximately 2.2 million new cases of lung cancer were reported globally, accounting for 11.4% of all newly diagnosed cancers. The number of deaths reached around 1.8 million, representing 18% of cancer-related fatalities.[1,2] In China, lung cancer ranks first among malignant tumors, with over 800,000 new cases and around 700,000 deaths annually. The risk of lung cancer is higher in men than in women, with an incidence rate of approximately 50 per 100,000 in men and 20 per 100,000 in women. Smoking is the primary risk factor, linked to 80% to 90% of lung cancer cases. Other environmental factors, including secondhand smoke, air pollution (PM2.5), and occupational exposures (e.g., asbestos), also significantly increase the risk. The 5-year survival rate for early-stage lung cancer is about 60% to 80%, while for late-stage (stage IV) lung cancer, it drops to only 5% to 10%.[3–6]

In recent years, studies have shown that air pollution, especially fine particulate matter (PM2.5), is closely associated with lung cancer. Air pollutants may promote the onset and progression of lung cancer through mechanisms such as inducing DNA damage, triggering inflammatory responses, and altering gene expression. Additionally, there is an interaction between air pollution and genetic susceptibility, with individuals at high genetic risk facing significantly increased lung cancer risk when exposed to air pollution. These findings suggest that air pollution is a critical environmental factor influencing lung cancer incidence, emphasizing the need for public health strategies to reduce exposure and protect high-risk populations.[7–10]

Mendelian randomization (MR) studies have played a crucial role in exploring the causal relationships of lung cancer risk factors. By using genetic variations as instrumental variables (IVs), MR studies effectively eliminate confounding factors and reverse causality, revealing causal links between various factors and lung cancer.[11–13]

The advantage of combining MR with meta-analysis lies in its ability to integrate findings from multiple studies, reduce bias, explore heterogeneity, and significantly enhance the generalizability and reliability of research conclusions. By incorporating data from different studies, this approach enables a more comprehensive analysis of the relationship between exposure and outcomes, improves statistical power, and allows for more precise evaluation of causal relationships with small effect sizes. This method provides more robust and credible conclusions, deepens the understanding of complex research questions, and offers stronger evidence to support decision-making in public health and clinical practice.

However, there has been no comprehensive research on MR combined with meta-analysis to investigate the causal relationship between air pollution indices and lung cancer. This study employs MR combined with meta-analysis to explore the causal relationships between specific air pollution indices and lung cancer, providing more precise theoretical support for clinical practice and research.

2. Methods and materials

2.1. Study design

In the research process, exposure and outcome data were first collected and preprocessed. Subsequently, the preprocessed exposure data, consisting of 5 air pollution indicators, were analyzed through Mendelian randomization (MR) with lung cancer data from 2 additional databases. Meta-analysis was then conducted on the inverse-variance weighted (IVW) results from the 2 MR analyses, followed by multiple corrections to the meta-analysis results to ensure data accuracy.

Finally, reverse MR analysis was performed on the positive air pollution indicators that showed significant causal associations with lung cancer, enabling a more precise understanding of the causal relationship between the 2.[14,15] In addition, a flow chart was drawn (Fig. 1).

2.2. Exposure data: genome-wide association studies for 5 air pollution indices data sources

The 5 air pollution indices in this study were sourced from the publicly available open Genome-wide association studies (GWAS) database, and all the data originated from the UK Biobank. Additionally, these air pollution indices were studied exclusively in European populations (Table 1).

The air pollution data download link is:

ukb-b-12417: https://gwas.mrcieu.ac.uk/files/ukb-b-589/ukb-b-589.vcf.gz
ukb-b-2618: https://gwas.mrcieu.ac.uk/files/ukb-b-2618/ukb-b-2618.vcf.gz
ukb-b-10817: https://gwas.mrcieu.ac.uk/files/ukb-b-10817/ukb-b-10817.vcf.gz

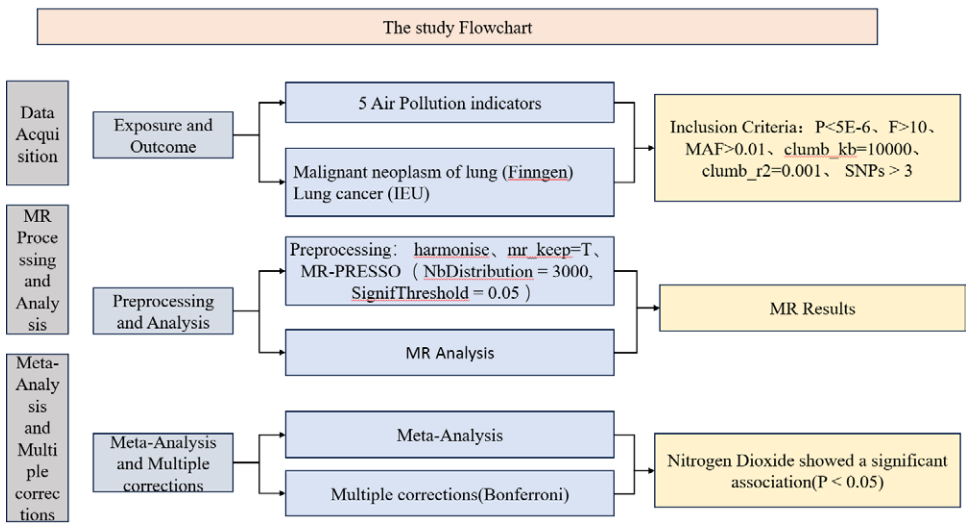

Figure 1. The process flowchart of the research methodology.

Table 1

Air pollution data.

GWAS ID	Trait	Sample size	SNPs	Consortium	Population
ukb-b-12417	Nitrogen oxides	456,380	9,851,867	MRC-IEU	European
ukb-b-2618	Nitrogen dioxide	456,180	9,851,867	MRC-IEU	European
ukb-b-10817	PM2.5 μm	423,796	9,851,867	MRC-IEU	European
ukb-b-12963	PM2.5 to 10 μm	423,796	9,851,867	MRC-IEU	European
ukb-b-589	PM10 μm	455,314	9,851,867	MRC-IEU	European

GWAS = genome-wide association study, SNPs = single nucleotide polymorphisms.

ukb-b-12963: https://gwas.mrcieu.ac.uk/files/ukb-b-12963/ukb-b-12963.vcf.gz
ukb-b-589: https://gwas.mrcieu.ac.uk/files/ukb-b-589/ukb-b-589.vcf.gz

According to recent research, we set the single nucleotide polymorphism (SNP) selection criteria for the 5 types of air pollution data to $P < 5 \times 10^{-6}$, $F > 10$, minor allele frequency (MAF) > 0.01, clump_kb = 10,000, and clump_r^2 = 0.001. After applying these criteria, 379 SNPs met the standards.[16–18]

2.3. Genome-wide association study data sources for lung cancer

The study accounted for the diversity of data sources. To avoid overlap between air pollution and lung cancer data, lung cancer outcome data were obtained from 2 different databases. The first set of lung cancer data was sourced from the latest Finngen R12 database, comprising 9639 cases and 378,749 controls, with 21,325,248 SNPs analyzed.[19] The download link is: gs://finngen-public-data-r12/summary_stats/release/finngen_R12_C3_BRONCHUS_LUNG_EXALLC.gz. The second set of lung cancer data was obtained from the OpenGWAS database, originating from the TRICL consortium, with a GWAS identifier of ieu-a-987. The download link is: https://gwas.mrcieu.ac.uk/files/ieu-a-987/ieu-a-987.vcf.gz. This dataset included 29,863 cases and 55,586 controls, with 10,439,018 SNPs analyzed. Both datasets pertain to European populations.

The criteria for data selection were as follows: preference was given to datasets with larger sample sizes, broader representation of data source organizations, and consistency between the study population and the exposure data.

2.4. Criteria for selection of instrumental variables

In Mendelian randomization (MR) studies, selecting effective IVs is crucial. First, this study adopted a selection threshold of *P*-value less than 5×10^{-6} to ensure that only SNPs strongly associated with various air pollutants were retained. For all air pollution data, the number of associated SNPs exceeded 3, ensuring the representativeness and relevance of the data.

Next, to further filter strong IVs, the study calculated the *F*-statistic value for each SNP using the formula $F = (\text{beta/se})$.[2] Only SNPs with an *F*-value >10 were retained, a step that helps eliminate weak IVs and enhances the reliability of the study results (Table S1, Supplemental Digital Content, https://links.lww.com/MD/P36). And a total of 2095 SNPs with $F > 10$ were identified, all showing a strong association with the air pollution index. Additionally, the MAF was calculated using the effect allele frequency (eaf). If eaf was <0.5, MAF was set to eaf; otherwise, it was set to 1-eaf. Only SNPs with an MAF >0.01 were retained to exclude rare variants that might affect the study results.

Finally, the filtered data were formatted for MR analysis and subjected to linkage disequilibrium (LD) clumping to avoid the impact of LD on result accuracy. The specific criteria for this step were a distance threshold set at 10, 000 kilobase pairs (kb) and an LD threshold set at 0.001. These steps ensured the independence of the IVs and the precision of the results.[20–22]

3. Statistical analysis

3.1. Causal validation of the 5 types of air pollution to lung cancer

All data analyses in this study were conducted using R version 4.4.2 (https://www.r-project.org/). First, SNP data from lung cancer outcome data were matched with the 5 air pollution exposure indices. Palindromic SNPs in the data were processed based on the criterion action = 2. Additionally, only data where mr_keep = true were retained.

Before applying the MR-PRESSO method, horizontal pleiotropy was tested on the processed data. If an SNP had a *P*-value <.05, it was considered horizontally pleiotropic and identified as an outlier, which was then removed using the MR-PRESSO method to ensure data accuracy. The MR-PRESSO parameters were set to NbDistribution = 3000 and SignifThreshold = 0.05. SNPs with *P*-values >.05 were deemed free of outliers.[23–25]

After refining the data, a heterogeneity test was performed prior to Mendelian randomization (MR) analysis. Although data heterogeneity has a minimal impact on results, SNPs showing significant heterogeneity (Q_pval < 0.05) were analyzed using the IVW random-effects model, while those without significant heterogeneity were analyzed using the IVW fixed-effects model. Regardless of heterogeneity, all data were further analyzed using the MR-Egger and weighted median methods, and odds ratios were calculated.

To enhance the reliability of the results, meta-analysis was performed on the MR analysis results for air pollution and both sets of lung cancer data. The significance *P*-values from the meta-analysis were adjusted using Bonferroni correction to reduce the likelihood of Type I errors. When conducting multiple hypothesis tests, the risk of false positives increases, necessitating appropriate multiple testing correction. The Bonferroni correction is a widely used and straightforward method that adjusts the significance threshold by dividing the desired alpha level by the total number of tests. This approach effectively controls the family-wise error rate and ensures the overall validity of statistical results. It is commonly applied in large-scale studies such as GWAS and MR analyses, where numerous comparisons are performed simultaneously. Ultimately, only 1 air pollution indicator showed a significant association after MR analysis and multiple corrections.[14]

3.2. Reverse causality validation between the strongly significant causally associated air pollution index and lung cancer

In this process, the strongly significant causally associated air pollution data were used as the outcome data, while lung cancer was treated as exposure. The same instrumental variable selection and data analysis methods as those used in the forward analysis were applied. The primary aim of this step was

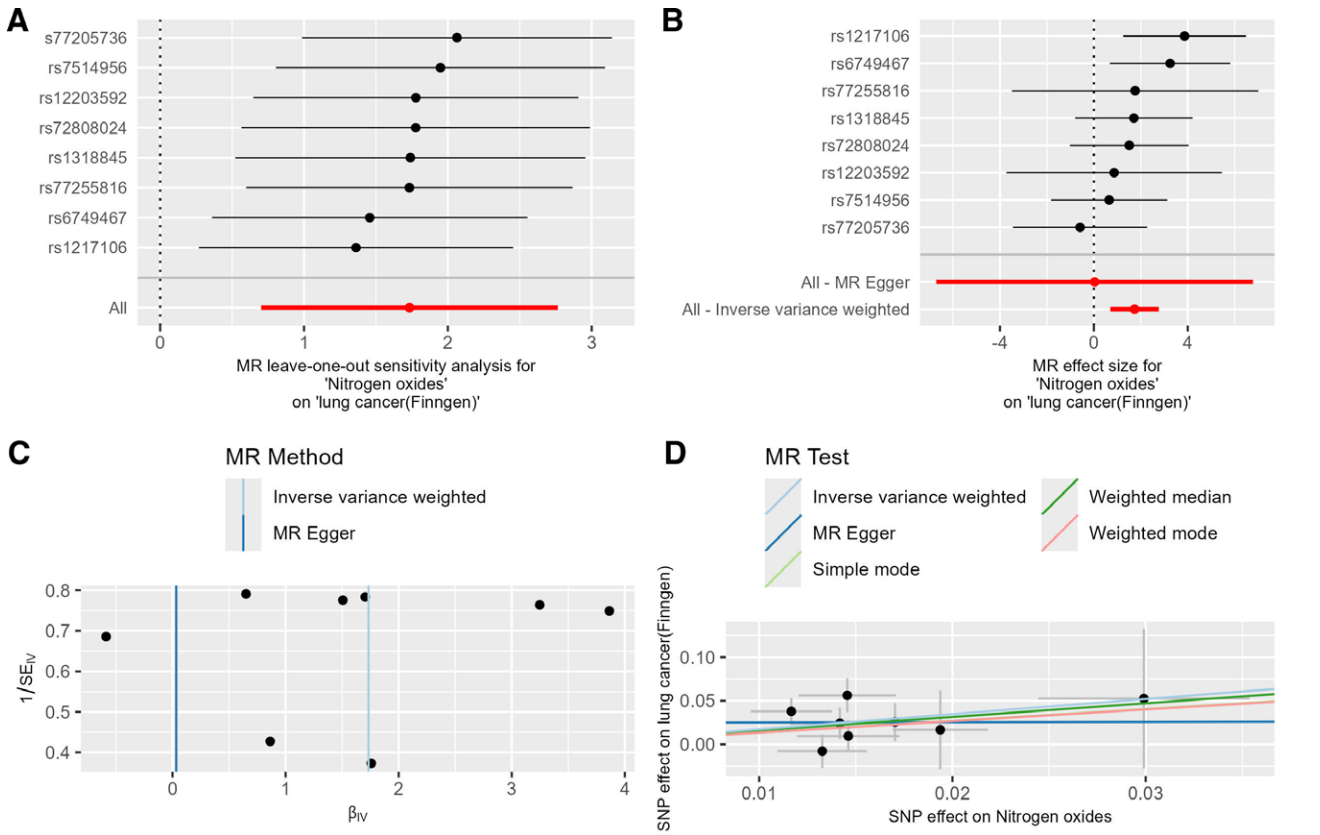

Figure 2. Combined MR plots of nitrogen dioxide on lung cancer (FinnGen). MR = Mendelian randomization.

to validate the directionality of the causal relationship between the 2. Therefore, the same thresholds and criteria were used to select valid IVs, and similar data processing and analysis were performed to determine the directionality of the relationship.

3.3. Sensitivity analysis

Horizontal pleiotropy means that different treatments or interventions may have different effects on different people or situations, and these effects may be mistakenly attributed to differences between the experimental and control groups rather than the actual treatment effect. To minimize the impact of horizontal pleiotropy on the experimental results, we performed a horizontal pleiotropy test on the GWAS data and removed outliers using MR-PRESSO for SNPs with significant horizontal pleiotropy (P-val < .05) (Table S2, Supplemental Digital Content, https://links.lww.com/MD/P36). The specific exclusion criteria were NbDistribution = 3000 and SignifThreshold = 0.05.[26,27]

Heterogeneity refers to the diversity or variability among research subjects, observations, or experimental conditions. In statistics and research methodology, heterogeneity typically refers to differences among samples or individuals, which may arise from individual characteristics, environmental conditions, or other factors. Heterogeneity is very common in research and can manifest in many ways, including physiological and psychological differences between individuals, variations in socioeconomic status, and environmental influences. This diversity and variability make research findings more generalizable and representative, but they also increase the complexity and difficulty of interpretation.

During the analysis process, we also conducted heterogeneity tests on the data. For SNPs with significant heterogeneity (Q_pval < .05), we used the IVW random-effects model for MR analysis. Otherwise, we used the fixed-effects model to ensure the accuracy and reliability of the results (Table S3, Supplemental Digital Content, https://links.lww.com/MD/P36).[28,29]

4. Results

4.1. Causal validation of the 5 types of air pollution to lung cancer

The 5 air pollution indices were analyzed through MR analysis with 2 sets of lung cancer data. Meta-analysis was conducted on the IVW results from the MR analyses, and multiple corrections were applied to the thresholds after the meta-analysis. Ultimately, only nitrogen dioxide (NO_2) showed a significant association with lung cancer.

Specifically, the MR analysis of nitrogen dioxide with lung cancer data from the Finngen R12 database yielded an IVW result with an OR of 5.660 (95% CI: 2.065–15.513, $P = 7.522 \times 10^{-4}$). The direction of the β values was consistent across the IVW, MR-Egger, and weighted median methods, indicating that NO_2 is a risk factor for lung cancer. Additionally, an MR scatter plot was created for this result (Fig. 2).

Similarly, the MR analysis with lung cancer data from the OpenGWAS database produced an OR of 2.635 (95% CI: 1.271–5.463, $P = 9.218 \times 10^{-3}$). The direction of the β values was also consistent across all 3 methods, further confirming NO_2 as a risk factor for lung cancer. A scatter plot was also generated for this analysis (Fig. 3) (Table S4, Supplemental Digital Content, https://links.lww.com/MD/P36).

Based on the 2 MR analyses of air pollution and lung cancer, we conducted a meta-analysis of the most significant result, the IVW result, from the MR analysis (Table S5, Supplemental Digital Content, https://links.lww.com/MD/P36). We then performed multiple corrections on the significant P-value from the meta-analysis using the Bonferroni method. The specific result was an OR value of 3.426 (95% CI: 1.897–6.186, $P = 2.21 \times 10^{-4}$)

(Table S6, Supplemental Digital Content, https://links.lww.com/MD/P36). Additionally, we created a forest plot for this meta-analysis result (Fig. 4). The evidence indicates that Nitrogen dioxide shows a significant association with lung cancer after combining MR analysis, meta-analysis, and multiple corrections.

The OR differences between the 2 databases may stem from population heterogeneity, sample size, and differences in exposure measurement methods. FinnGen R12 is based on the Finnish population, whose genetic background or environmental exposures (such as regional high pollution levels or smoking rates) may enhance the effect of NO_2, leading to a higher OR value (5.66). In contrast, OpenGWAS includes a broader population, and exposure assessments may be based on global models. Individual exposure classification errors could dilute the effect, resulting in a lower OR value (2.64). Additionally, the smaller sample size in FinnGen may lead to the "winner's curse" (overestimating the effect), while the larger sample size in OpenGWAS provides more stable but conservative estimates. Differences in the definition of lung cancer subtypes or diagnostic criteria (e.g., whether histological types are distinguished) between the 2 studies may also affect the results. Although the OR values differ, both methods show consistent β value directions, and statistical significance (FinnGen $P = 7.5 \times 10^{-4}$ vs OpenGWAS's $P = 9.2 \times 10^{-3}$) supports a causal association between NO_2 and lung cancer. This difference highlights the importance of population specificity and exposure measurement accuracy in MR studies, but the consistency of the conclusions across datasets provides evidence for the carcinogenic role of NO_2. Future research should combine refined exposure assessments, multi-population validation, and mechanistic studies to further elucidate the biological basis of the effect differences.

4.2. Reverse causality validation between the strongly significant causally associated air pollution index and lung cancer

In the reverse MR validation process, no significant association was found between nitrogen dioxide (NO_2) and lung

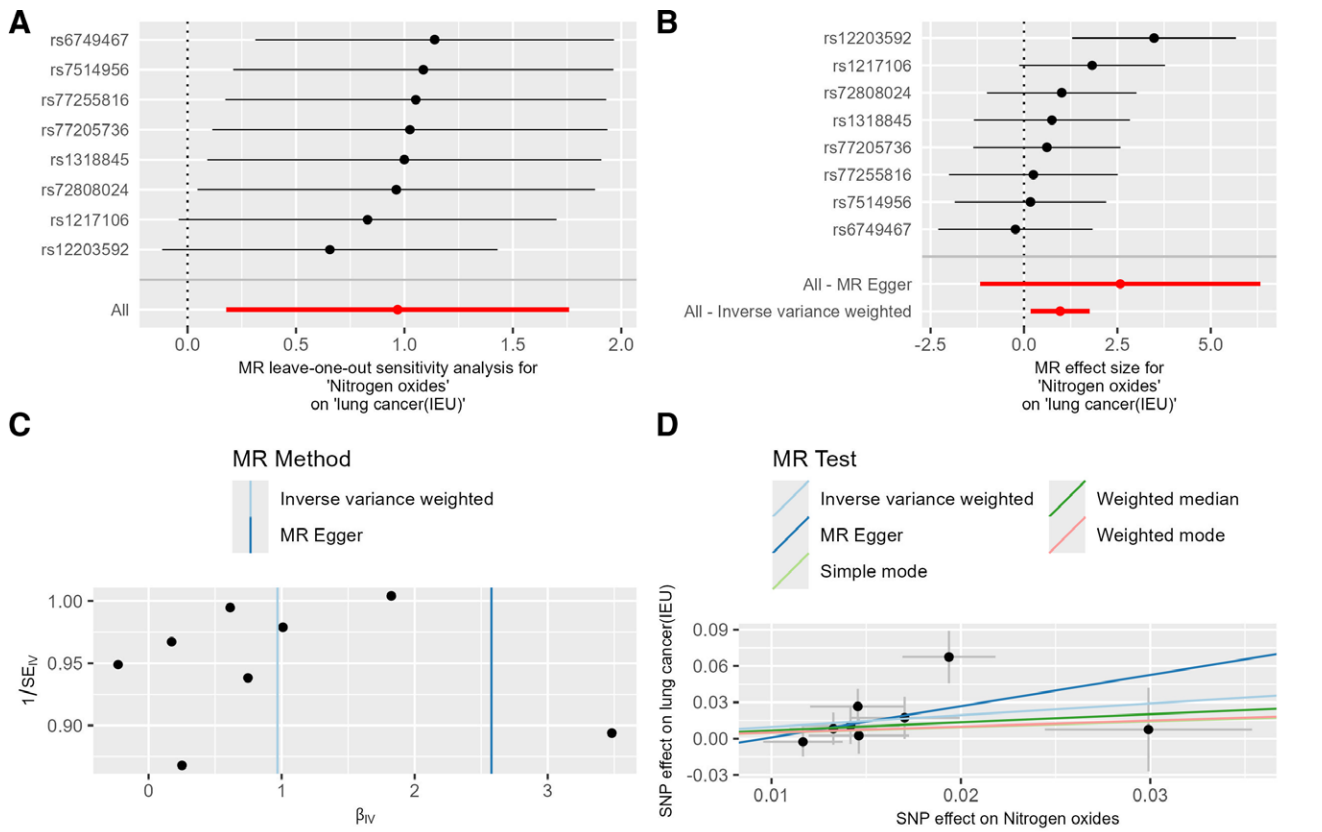

Figure 3. Combined MR plots of nitrogen dioxide on lung cancer (openGWAS). GWAS = genome-wide association study, MR = Mendelian randomization.

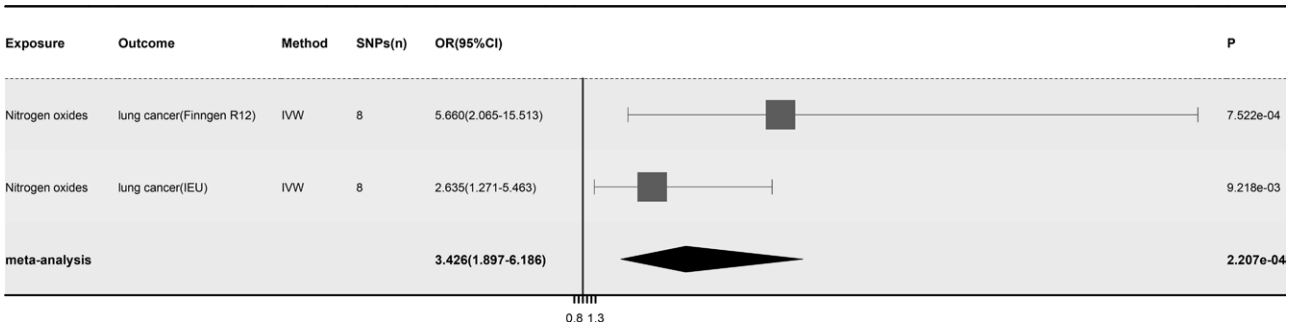

Figure 4. Forest plot of nitrogen dioxide after meta-analysis.

cancer. Specifically, when lung cancer data from the Finngen R12 database were used as the exposure, the OR was 1.000 (95% CI: 0.992–1.008, $P = .969$). Similarly, when lung cancer data from the OpenGWAS database were used as the exposure, the OR was 1.004 (95% CI: 0.995–1.012, $P = .372$). In summary, these results indicate that nitrogen dioxide showed no evidence of reverse causality with lung cancer in either dataset (Table S7, Supplemental Digital Content, https://links.lww.com/MD/P36).

5. Discussion

The study explored the causal relationship between 5 air pollution indices and lung cancer using bidirectional MR combined with meta-analysis and multiple corrections. The results revealed that only nitrogen dioxide (NO_2) showed a strong significant causal association with lung cancer, and no reverse causal relationship was found between them.

As a major air pollutant, NO_2 is closely associated with the incidence and mortality of various cancers. Epidemiological studies suggest that long-term exposure to NO_2 significantly increases the risk of cancers such as breast, ovarian, uterine, lung, bladder, and colorectal cancer. NO_2 promotes tumor initiation and progression through mechanisms such as inducing oxidative stress and chronic inflammation, causing DNA damage, and interfering with hormone levels. However, the causal relationship between NO_2 and cancer has not been fully clarified. This study, from a genetic perspective, confirmed a significant association between NO_2 and lung cancer, identifying NO_2 as a potential risk factor for lung cancer while confirming the absence of reverse causality.[30–32]

Nitrogen dioxide is a common air pollutant with the chemical formula NO_2 and a molecular weight of 46.0055 g/mol. At room temperature, it appears as a reddish-brown gas with a pungent odor and dissolves readily in water to form nitric acid and nitrous acid. Its sources include natural phenomena such as volcanic activity, forest fires, and bacterial action, as well as human activities such as fossil fuel combustion, vehicle emissions, power plant operations, and industrial discharges. Long-term exposure to NO_2 can lead to respiratory diseases (e.g., asthma, bronchitis, chronic obstructive pulmonary disease), cardiovascular diseases (e.g., heart attacks and strokes), increased cancer risk, and suppressed immune function. Environmentally, NO_2 contributes to acid rain, causing soil and water acidification and harming ecosystems. It also reacts with other pollutants under sunlight to form ozone, a key component of photochemical smog. Controlling NO_2 emissions is therefore crucial for improving air quality, protecting human health, and preserving ecosystems.[32–34]

In recent years, research into the relationship between NO_2 and lung cancer has advanced significantly. Epidemiological studies have shown a strong correlation between long-term exposure to NO_2 and increased lung cancer risk, particularly lung adenocarcinoma. NO_2, as a critical air pollutant, promotes lung cancer initiation and progression through mechanisms such as oxidative stress, chronic inflammation, DNA integrity disruption, and interference with cellular signaling pathways. These mechanisms exacerbate carcinogenesis through direct lung tissue damage and indirect genetic mutations. Additionally, long-term NO_2 exposure may act synergistically with smoking, occupational exposure, and other environmental pollutants, further elevating lung cancer incidence. Recent Mendelian randomization (MR) studies, using genetic IVs, have preliminarily validated the causal relationship between NO_2 exposure and lung cancer, suggesting NO_2 as a significant environmental risk factor, especially among populations with long-term exposure.[35–37]

Nitrogen dioxide contributes to lung cancer development through multiple mechanisms. First, NO_2 exposure induces oxidative stress, leading to increased reactive oxygen species levels, which damage DNA, proteins, and lipids, activating tumor-promoting signaling pathways. Second, NO_2 induces chronic inflammation in the lungs, elevating pro-inflammatory factors such as IL-6 and TNF-α, which create conditions conducive to cancer cell proliferation and invasion while suppressing antitumor immune responses, fostering a tumor-promoting microenvironment. Additionally, NO_2 may directly or indirectly cause DNA damage, genetic mutations, and chromosomal instability, activating oncogenes such as KRAS and EGFR or inactivating tumor suppressor genes such as TP53.[38,39] NO_2 also disrupts mitochondrial function, impairing the balance between apoptosis and proliferation, enhancing cancer cell survival, and promoting abnormal cell accumulation. Epigenetic mechanisms are also affected, with NO_2 regulating DNA methylation, histone modification, and noncoding RNA, altering gene expression patterns and driving tumorigenesis. Moreover, NO_2 interacts with other air pollutants like PM2.5 and ozone, exacerbating oxidative stress and inflammation, thereby synergistically promoting lung cancer development. These mechanisms provide significant biological evidence for NO_2's role in lung cancer pathogenesis.[40–43]

Based on these findings, various strategies could mitigate NO_2's impact on lung cancer patients. First, antioxidants such as vitamins C and E can neutralize NO_2-induced oxidative stress, reducing cellular damage. Second, anti-inflammatory drugs, such as nonsteroidal anti-inflammatory drugs (NSAIDs), can alleviate chronic inflammation and improve patient outcomes. Immunotherapies, including immune checkpoint inhibitors like PD-1/PD-L1 inhibitors, may restore the immune system's ability to attack cancer cells, while gene repair drugs like PARP inhibitors can block DNA repair pathways in cancer cells, reducing NO_2-induced genetic damage. Additionally, non-intubated thoracoscopic surgery, a minimally invasive procedure, can directly remove lesions or collect samples without general anesthesia, significantly reducing postoperative complications and accelerating recovery, offering essential support for personalized treatment. Environmentally, reducing industrial and transportation emissions and increasing green spaces to lower atmospheric NO_2 concentrations can effectively mitigate its negative effects on patients. These combined measures work synergistically to improve treatment outcomes and quality of life for lung cancer patients.

This study holds significant public health and medical implications. It highlights the critical role of reducing NO_2 concentrations in lung cancer prevention, advocating for stricter air quality management policies to minimize health risks from air pollution. Furthermore, exploring the causal association between NO_2 and lung cancer enhances the health risk assessment system for air pollution, aiding in identifying high-risk populations and implementing targeted preventive measures. Revealing the relationship between air pollution and lung cancer also provides a solid foundation for future research on the impact of air pollution on other cancer types, enabling a more comprehensive understanding of its health effects. Overall, studying the link between NO_2 and lung cancer not only supports the implementation of stricter air quality standards but also offers new insights into early prevention and intervention strategies for high-risk populations. Reducing NO_2 levels in the air could significantly improve public health, decrease lung cancer incidence, and foster healthier living environments.

Building on multiple observational studies, this study genetically validated the causal association between air pollution and lung cancer, achieving a fully randomized genetic-controlled trial. This approach effectively mitigates the confounding factors common in observational studies, providing a more precise understanding of the relationship between air pollution and lung cancer risk. By combining Mendelian randomization (MR) analysis and meta-analysis, the study results demonstrate higher reliability and scientific validity. However, certain limitations exist. Due to data constraints, the study primarily focused on

European populations, and its findings may not be fully applicable to diverse global populations. Future research should expand to include other ethnic groups and regions to further validate and complement these findings. Despite these limitations, this study provides critical evidence for public health policymaking and a deeper understanding of the risk of air pollution in lung cancer. These findings not only support the implementation of stricter air quality standards but also offer novel approaches to early prevention and intervention for high-risk populations. Lowering atmospheric NO_2 concentrations could significantly enhance public health, reduce lung cancer incidence, and promote healthier living environments.

6. Conclusions

This study demonstrates a significant causal association between nitrogen dioxide (NO_2) and lung cancer, indicating that NO_2 may be a potential risk factor for lung cancer.

Acknowledgments

Firstly, we express our profound thanks to all individuals and researchers who participated in the GWAS data for this research. Additionally, we extend our sincere gratitude and respect to the personnel involved with the associated public databases. Lastly, our heartfelt appreciation goes out to every author who played a role in contributing to this study.

Author contributions

Conceptualization: Xiaomin Wang, Guihua Xiao, Wanxian Xu, Changguo Ni.
Data curation: Xiaomin Wang, Guihua Xiao.
Formal analysis: Xiaomin Wang, Guihua Xiao, Wanxian Xu, Changguo Ni.
Investigation: Xiaomin Wang, Guihua Xiao.
Methodology: Xiaomin Wang, Guihua Xiao, Wanxian Xu, Changguo Ni.
Project administration: Xiaomin Wang, Guihua Xiao, Wanxian Xu, Changguo Ni.
Resources: Xiaomin Wang, Changguo Ni.
Software: Xiaomin Wang, Wanxian Xu.
Supervision: Changguo Ni.
Validation: Xiaomin Wang, Wanxian Xu, Changguo Ni.
Visualization: Xiaomin Wang, Guihua Xiao, Wanxian Xu.
Writing – original draft: Xiaomin Wang, Guihua Xiao.
Writing – review & editing: Wanxian Xu, Changguo Ni.

References

[1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
[2] Barlesi F, Dixmier A, Debieuvre D, et al. Effectiveness and safety of nivolumab in the treatment of lung cancer patients in France: preliminary results from the real-world EVIDENS study. Oncoimmunology. 2020;9:1744898.
[3] Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet. 2021;398:535–54.
[4] Li Y, Wu X, Yang P, Jiang G, Luo Y. Machine learning for lung cancer diagnosis, treatment, and prognosis. Genomics Proteomics Bioinformatics. 2022;20:850–66.
[5] Detterbeck FC, Woodard GA, Bader AS, et al. The proposed ninth edition TNM classification of lung cancer. Chest. 2024;166:882–95.
[6] Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: current application and future perspective. Semin Cancer Biol. 2023;89:30–7.
[7] Chowdhury S, Pillarisetti A, Oberholzer A, et al. A global review of the state of the evidence of household air pollution's contribution to ambient fine particulate matter and their related health impacts. Environ Int. 2023;173:107835.
[8] Rahmioglu N, Mortlock S, Ghiasi M, et al; DBDS Genomic Consortium. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. Nat Genet. 2023;55:423–36.
[9] Kong J, Yan S, Cao X, et al. Quantitative source apportionment and health risk assessment for polycyclic aromatic hydrocarbon and their derivatives in indoor dust from housing and public buildings of a megacity in China. J Hazard Mater. 2024;486:137057.
[10] Lugg ST, Scott A, Parekh D, Naidu B, Thickett DR. Cigarette smoke exposure and alveolar macrophages: mechanisms for lung disease. Thorax. 2022;77:94–101.
[11] Bouras E, Karhunen V, Gill D, et al; PRACTICAL consortium. Circulating inflammatory cytokines and risk of five cancers: a Mendelian randomization analysis. BMC Med. 2022;20:3.
[12] Larsson SC, Butterworth AS, Burgess S. Mendelian randomization for cardiovascular diseases: principles and applications. Eur Heart J. 2023;44:4913–24.
[13] Li J, Tang M, Gao X, Tian S, Liu W. Mendelian randomization analyses explore the relationship between cathepsins and lung cancer. Commun Biol. 2023;6:1019.
[14] Luo J, le Cessie S, van Heemst D, Noordam R. Diet-derived circulating antioxidants and risk of coronary heart disease: a mendelian randomization study. J Am Coll Cardiol. 2021;77:45–54.
[15] Jiang F, Zhao J, Sun J, et al. Impact of ambient air pollution on colorectal cancer risk and survival: insights from a prospective cohort and epigenetic Mendelian randomization study. EBioMedicine. 2024;103:105126.
[16] Shen J, Guo Y, Cao R. The relationship between amino acids and gastroesophageal reflux disease: evidence from a mendelian randomization analysis combined with a meta-analysis. Front Immunol. 2025;16:1420132.
[17] Zhu W, Fu M, Li Q, et al. Amino acid metabolism-related genes as potential biomarkers and the role of MATN3 in stomach adenocarcinoma: a bioinformatics, mendelian randomization and experimental validation study. Int Immunopharmacol. 2024;143:113253.
[18] Xie F, Feng Z, Xu B. Metabolic characteristics of gut microbiota and insomnia: evidence from a mendelian randomization analysis. Nutrients. 2024;16:2943.
[19] Kurki MI, Karjalainen J, Palta P, et al; FinnGen. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613:508–18.
[20] Vaucher J, Keating BJ, Lasserre AM, et al. Cannabis use and risk of schizophrenia: a Mendelian randomization study. Mol Psychiatry. 2018;23:1287–92.
[21] Papadopoulou A, Åsvold BO, Burgess S, et al. Height, autoimmune thyroid disease, and thyroid cancer: a mendelian randomization study. Thyroid. 2023;33:1476–82.
[22] Kintu C, Soremekun O, Kamiza AB, et al. The causal effects of lipid traits on kidney function in Africans: bidirectional and multivariable Mendelian-randomization study. EBioMedicine. 2023;90:104537.
[23] Fang P, Liu X, Qiu Y, et al. Exploring causal correlations between inflammatory cytokines and ankylosing spondylitis: a bidirectional mendelian-randomization study. Front Immunol. 2023;14:1285106.
[24] Long Y, Tang L, Zhou Y, Zhao S, Zhu H. Causal relationship between gut microbiota and cancers: a two-sample Mendelian randomisation study. BMC Med. 2023;21:66.
[25] Yang H, Shi P, Li M, et al. Mendelian-randomization study reveals causal relationships between nitrogen dioxide and gut microbiota. Ecotoxicol Environ Saf. 2023;267:115660.
[26] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.
[27] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. Epidemiology (Cambridge). 2017;28:30–42.
[28] Mutie PM, Pomares-Millan H, Atabaki-Pasdar N, et al. Investigating the causal relationships between excess adiposity and cardiometabolic health in men and women. Diabetologia. 2023;66:321–35.
[29] Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46:1985–98.
[30] Raaschou-Nielsen O, Andersen ZJ, Beelen R, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol. 2013;14:813–22.
[31] Hystad P, Villeneuve PJ, Goldberg MS, Crouse DL, Johnson K; Canadian Cancer Registries Epidemiology Research Group. Exposure

to traffic-related air pollution and the risk of developing breast cancer among women in eight Canadian provinces: a case-control study. Environ Int. 2015;74:240–8.
[32] Shi L, Rosenberg A, Wang Y, et al. Low-concentration air pollution and mortality in american older adults: a national cohort analysis (2001-2017). Environ Sci Technol. 2022;56:7194–202.
[33] Lechner MG, Bernardo AC, Lampe A, Praw SS, Tam SH, Angell TE. Changes in stage distribution and disease-specific survival in differentiated thyroid cancer with transition to american joint committee on cancer 8th edition: a systematic review and meta-analysis. Oncologist. 2021;26:e251–60.
[34] Kansal A. Sources and reactivity of NMHCs and VOCs in the atmosphere: a review. J Hazard Mater. 2009;166:17–26.
[35] Li W, Wang W. Causal effects of exposure to ambient air pollution on cancer risk: Insights from genetic evidence. Sci Total Environ. 2024;912:168843.
[36] 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;50:693–8.
[37] Verbanck M, Chen CY, Neale B, Do R. Publisher correction: detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:1196.
[38] Quezada-Maldonado EM, Sánchez-Pérez Y, Chirino YI, García-Cuellar CM. Airborne particulate matter induces oxidative damage, DNA adduct formation and alterations in DNA repair pathways. Environ Pollut. 2021;287:117313.
[39] Li R, Zhou R, Zhang J. Function of PM2.5 in the pathogenesis of lung cancer and chronic airway inflammatory diseases. Oncol Lett. 2018;15:7506–14.
[40] Holme JA, Vondráček J, Machala M, et al. Lung cancer associated with combustion particles and fine particulate matter (PM(2.5)) – the roles of polycyclic aromatic hydrocarbons (PAHs) and the aryl hydrocarbon receptor (AhR). Biochem Pharmacol. 2023;216:115801.
[41] Chen CY, Huang KY, Chen CC, et al. The role of PM2.5 exposure in lung cancer: mechanisms, genetic factors, and clinical implications. EMBO Mol Med. 2025;17:31–40.
[42] Lakhdar R, Mumby S, Abubakar-Waziri H, Porter A, Adcock IM, Chung KF. Lung toxicity of particulates and gaseous pollutants using ex-vivo airway epithelial cell culture systems. Environ Pollut. 2022;305:119323.
[43] Leclercq B, Kluza J, Antherieu S, et al. Air pollution-derived PM(2.5) impairs mitochondrial function in healthy and chronic obstructive pulmonary diseased human bronchial epithelial cells. Environ Pollut. 2018;243:1434–49.