scientific reports

Check for updates

OPEN

Disentangling the effects of various risk factors and trends in lung cancer mortality

Xiu Wu[1]✉, Jinting Zhang[2]✉ & Yilin Yoshida[1]

Lung cancer is a leading cause of mortality in oncological classifications, yet the impact of various risk factors on lung cancer mortality (LCM) in non-smokers remains unclear. This study aims to weigh out the diverse impact of multiple risk factors on LCM rates and identify trends in LCM rates worldwide. We initially employed Random Forest Tree (RFT) and Gradient Boosting Regression (GBR) to identify common primary factors influencing LCM. After eliminating four common primary factors, a comparative analysis between partial and Pearson correlations was conducted to filter out significant factors in the correlations between risk factors and LCM rates across 204 countries from 2005 to 2019. The findings show that excluding the impacts of occupational exposure to arsenic, smoking, residential radon, occupational exposure to silica, occupational exposure to asbestos, high systolic blood pressure, secondhand smoke, child wasting, and alcohol use had a considerably greater impact on LCM than particular matter pollution (PM2.5). Furthermore, a Multiple Joinpoint Regression analysis identified increasing trends of LCM rates in the 142 countries (e.g., China and India); decreasing trends in 38 countries (e.g., Denmark and Norway), and stable trends in 24 countries (e.g., Sudan, Mali, and Australia). This research suggests that in addition to considering the effects of occupational exposure to arsenic, smoking, residential radon, and occupational exposure to silica on LCM rates, occupational exposure to asbestos, high systolic blood pressure, secondhand smoke, child wasting, and alcohol use should be considered in lung cancer prevention strategies, especially in countries with increasing trends of LCM rates.

Keywords Random forest tree, Gradient boosting regression, Partial correlation, Multiple joinpoint regression, Lung cancer mortality

Lung cancer is a leading cause of mortality among cancer patients. It is characterized by a lack of specific early symptoms, and symptoms may not appear even in advanced stages, making early detection based on clinical signs difficult. By the time lung cancer is diagnosed, it is typically in stage IV, with a 5-year survival rate of just 22%[1]. When it was first identified in the 20th century, lung cancer was a rare malignancy that originated in the lungs[2]. However, it quickly became one of the major public health burdens in the U.S. According to the American Cancer Society, an estimated 350 people die from lung cancer each day in the U.S., exceeding the combined number of deaths from breast, prostate, and pancreatic cancers. Lung cancer causes 2.5 times more deaths than colon cancer. In 2022, there were 609,360 lung cancer deaths, and in 2023, over 238,000 new cases were diagnosed, with more than 127,000 deaths reported in the U.S. This de facto counted for the largest proportion death of cancer deaths[3]. Similarly, lung cancer in China had a higher mortality rate among the elderly than in other countries. It was the leading cancer death among cancer lineage. There were an estimated 0.87 million new cases and 0.766 million deaths in 2023[4].

According to data from the Genomic Data Commons Data Portal of the National Cancer Institute, lung cancer has an overall survival rate of less than 20%. 52.57% of lung cancer patients are female, and 69.72% of cases involve adenomas and adenocarcinomas (https://portal.gdc.cancer.gov/). In light of cancer over time statistic of lung cancer mortality (LCM) in age [0–84] from 1992 to 2014 in International Agency for Research Cancer, existed age-standardized rate (ASR) of male steeply declined from 43.7 to 26.7 per 100,000 people while the rate for female almost stagnated during 12.5–12.9 per 100,000 people over the same period, owing to the tobacco control. The crude rate per 100,000 of male decreased from 53.3 to 46.4, while that rate for female increased from 18.9 to 25. The cumulative risk (in percentage) for men decreased from 10 to 7, while that for women fluctuated from 2.8 to 3.1 (https://gco.iarc.fr/overtime/en/dataviz/). Based on a 2020–2021 cancer report

[1]School of Medicine, Tulane University, New Orleans, LA 70112, USA. [2]School of Resource and Environmental Science, Wuhan University, Wuhan 430070, China. ✉email: xwu17@tulane.edu; whuzjt@whu.edu.cn

from IARC, lung cancer was responsible for about 2.2 million deaths worldwide, accounting for approximately 11.4% of cancer-related deaths.

The question of how to accurately assess the multifaceted impact of risk factors on LCM rates remains unresolved. While smoking is well-established as the dominant risk factor[3,5,6], other factors such as occupational exposures[7–9], environmental pollutants[10–12], socioeconomic status[13,14], and comorbidities[15] have not been fully elucidated. Research has often relied on limited regional samples or focused on individual factors in isolation, leaving gaps in our understanding of the broader, multifaceted nature of lung cancer risk. Addressing these gaps are crucial, as the complex interplay of various factors may have a substantial impact on LCM. Clarifying these interconnections provided a more comprehensive understanding of how different risk factors contribute to LCM. In light of this, our research aims to provide a comprehensive evaluation of lung cancer risk factors, assessing the broader effects on LCM rates, including indirect and previously overshadowed factors. We also investigate the spatio-temporal disparities in LCM globally, offering new insights into the complexity of risk factors beyond smoking.

Data and method

We collected data on LCM and mortality associated with 73 risk factors in 204 countries for the period 2005 to 2019 from open datasets published by the Global Burden of Disease (https://vizhub.healthdata.org/gbd-results/). The dataset contained 3060 observations. Mortality associated with a risk factor is simply the estimated number of deaths associated with each of the 73 risk factors. The 73 risk factors are listed in Table 1.

Investigation Procedure

We conducted an analysis of LCM rates across 204 countries, as illustrated in Figure 1. The age-adjusted LCM rates were classified into five categories—very low risk, low risk, medium risk, high risk, and very high risk—based on a quintile classification.

First, Random Forest Tree (RFT) modeling and Gradient Boosting Regression (GBR) were employed to identify four major common risk factors among the 73 factors analyzed.

Second, partial correlation and bivariate Pearson correlation analyses were performed on the remaining factors.

Third, we compared the coefficient differences between partial correlation and bivariate Pearson correlation to filter out significantly impactful variables related to LCM rates.

Fourth, a GBR model was developed for the selected variables.

Finally, a Multiple Joinpoint Regression model was employed to reveal trends in LCM rates worldwide. The analysis framework is depicted in Figure 2.

Partial correlation analysis

Partial Correlation is used to analyze the relationship between two continuous variables while accounting for the influence of control variables. In this analysis, four control variables were included: occupational exposure to arsenic, smoking, residential radon, and occupational exposure to silica. The analysis begins by checking whether the LCM rates meet normality assumptions. As shown in Fig. 3(a), the LCM rates distribution is skewed. Therefore, a logarithmic transformation was applied to normalize the distribution, as illustrated in Fig. 3(b). This method is particularly useful for uncovering spurious or hidden correlations that may not be initially apparent[16].

Calculate the correlation coefficient r.

$$r = \frac{r_{XY} - r_{XY} r_{XZ}}{\sqrt{(1 - r_{YZ}^2)(1 - r_{XZ}^2)}} \tag{1}$$

Where r_{XY} is the coefficient of the relationship between X and Y.

r_{XZ} is the coefficient of the relationship between X and Z.

r_{YZ} is the coefficient of the relationship between Y and Z.

We made hypothesis below.

H_0: 73 risk factors were not related with LCM rates after the impacts of occupational exposure to arsenic, smoking, residential radon, occupational exposure to silica.

H_1: 73 risk factors were related with LCM after the impacts of occupational exposure to arsenic, smoking, residential radon, occupational exposure to silica.

Conduct t-test using the following.

$$t = r\sqrt{\frac{n - q - 2}{1 - r^2}} \tag{2}$$

where n is the sampling size, q refers to the order of partial correlations, $t \sim T_{df}$ = n-q-2.

calculate p-value and get the conclusion. If p value is less than 0.05, reject H_0, meaning 73 risk factors were significantly related to LCM rates.

Gradient Boosting Regression (GBR)

The GBR is a machine learning approach that develops a predictive model by combining an ensemble of weak learners in a sequential order. The rationale of gradient boosting is to interactively optimize a loss function by adding weak models to ensemble. Each weak model is trained to correct the previous errors sequentially. This method improves prediction powers and creates a highly accurate and robust predictive model that can handle

NO	Variable Name	Description
1	Ambient ozone pollution	Mortality rate from ozone of air pollution
2	Chewing tobacco	Mortality rate from tobacco
3	Child and maternal malnutrition	Mortality rate from child and maternal malnutrition
4	Child growth failure	Mortality rate from child growth failure
5	Child stunting	Mortality rate from child stunting
6	Child underweight	Mortality rate from child underweight
7	Child wasting	Mortality rate from child wasting
8	Childhood sexual abuse	Mortality rate from childhood sexual abuse
9	High body-mass index	Mortality rate from high body-mass index
10	High fasting plasma glucose	Mortality rate from high fasting plasma glucose
11	High LDL cholesterol	Mortality rate from high LDL cholesterol
12	High systolic blood pressure	Mortality rate from high systolic blood pressure
13	High temperature	Mortality rate from high temperature
14	Particulate matter pollution	Mortality rate from particulate matter pollution
15	Household air pollution	Mortality rate from household air pollution
16	Intimate partner violence	Mortality rate from intimate partner violence
17	No access to handwashing facility	Mortality rate from no access to handwashing facility
18	Non-exclusive breastfeeding	Mortality rate from non-exclusive breastfeeding
19	Residential radon	Mortality rate from residential radon
20	Secondhand smoke	Mortality rate from secondhand smoke
21	Unsafe sanitation	Mortality rate from unsafe sanitation
22	Low temperature	Mortality rate from low temperature
23	Unsafe water source	Mortality rate from unsafe water source
24	Smoking	Mortality rate from smoking
25	Diet high in processed meat	Mortality rate from diet high in processed meat
26	Diet high in red meat	Mortality rate from diet high in red meat
27	Diet high in sodium	Mortality rate from diet high in sodium
28	Diet high in sugar-sweetened beverages	Mortality rate from diet high in sugar-sweetened beverages
29	Diet high in trans fatty acids	Mortality rate from diet high in trans fatty acids
30	Diet low in calcium	Mortality rate from diet low in calcium
31	Diet low in fiber	Mortality rate from diet low in fiber
32	Diet low in fruits	Mortality rate from diet low in fruits
33	Diet low in legumes	Mortality rate from diet low in legumes
34	Diet low in milk	Mortality rate from diet low in milk
35	Diet low in nuts and seeds	Mortality rate from diet low in nuts and seeds
36	Diet low in polyunsaturated fatty acids	Mortality rate from diet low in polyunsaturated fatty acids
37	Diet low in seafood omega-3 fatty acids	Mortality rate from diet low in seafood omega-3 fatty acids
38	Diet low in vegetables	Mortality rate from diet low in vegetables
39	Diet low in whole grains	Mortality rate from diet low in whole grains
40	Dietary risks	Mortality rate from dietary risks
41	Discontinued breastfeeding	Mortality rate from discontinued breastfeeding
42	Drug use	Mortality rate from drug use
43	Iron deficiency	Mortality rate from iron deficiency
44	Kidney dysfunction	Mortality rate from kidney dysfunction
45	Unsafe sex	Mortality rate from unsafe sex
46	Short gestation	Mortality rate from short gestation
47	Suboptimal breastfeeding	Mortality rate from suboptimal breastfeeding
48	Low physical activity	Mortality rate from low physical activity
49	Alcohol use	Mortality rate from alcohol use
50	Occupational asthmagens	Mortality rate from occupational asthmagens
51	Occupational carcinogens	Mortality rate from occupational carcinogens
52	Occupational exposure to arsenic	Mortality rate from occupational exposure to arsenic
53	Occupational exposure to asbestos	Mortality rate from occupational exposure to asbestos
54	Occupational exposure to benzene	Mortality rate from occupational exposure to benzene
55	Occupational exposure to beryllium	Mortality rate from occupational exposure to beryllium
56	Occupational exposure to cadmium	Mortality rate from occupational exposure to cadmium
Continued		

NO	Variable Name	Description
57	Occupational exposure to chromium	Mortality rate from occupational exposure to chromium
58	Occupational exposure to diesel engine exhaust	Mortality rate from occupational exposure to diesel engine exhaust
59	Occupational exposure to formaldehyde	Mortality rate from occupational exposure to formaldehyde
60	Occupational exposure to nickel	Mortality rate from occupational exposure to nickel
61	Occupational exposure to polycyclic aromatic hydrocarbons	Mortality rate from occupational exposure to polycyclic aromatic hydrocarbons
62	Occupational exposure to silica	Mortality rate from occupational exposure to silica
63	Occupational exposure to sulfuric acid	Mortality rate from occupational exposure to sulfuric acid
64	Occupational exposure to trichloroethylene	Mortality rate from occupational exposure to trichloroethylene
65	Occupational injuries	Mortality rate from occupational injuries
66	Occupational particulate matter, gases, and fumes	Mortality rate from occupational particulate matter, gases, and fumes
67	Lead exposure	Mortality rate from lead exposure
68	Low birth weight	Mortality rate from low birth weight
69	Low birth weight and short gestation	Mortality rate from low birth weight and short gestation
70	Vitamin A deficiency	Mortality rate from vitamin A deficiency
71	Zinc deficiency	Mortality rate from zinc deficiency
72	Metabolic risks	Mortality rate from metabolic risks
73	Low bone mineral density	Mortality rate from low bone mineral density

Table 1. List of risk factors of a county used as independent variables in the analyses.

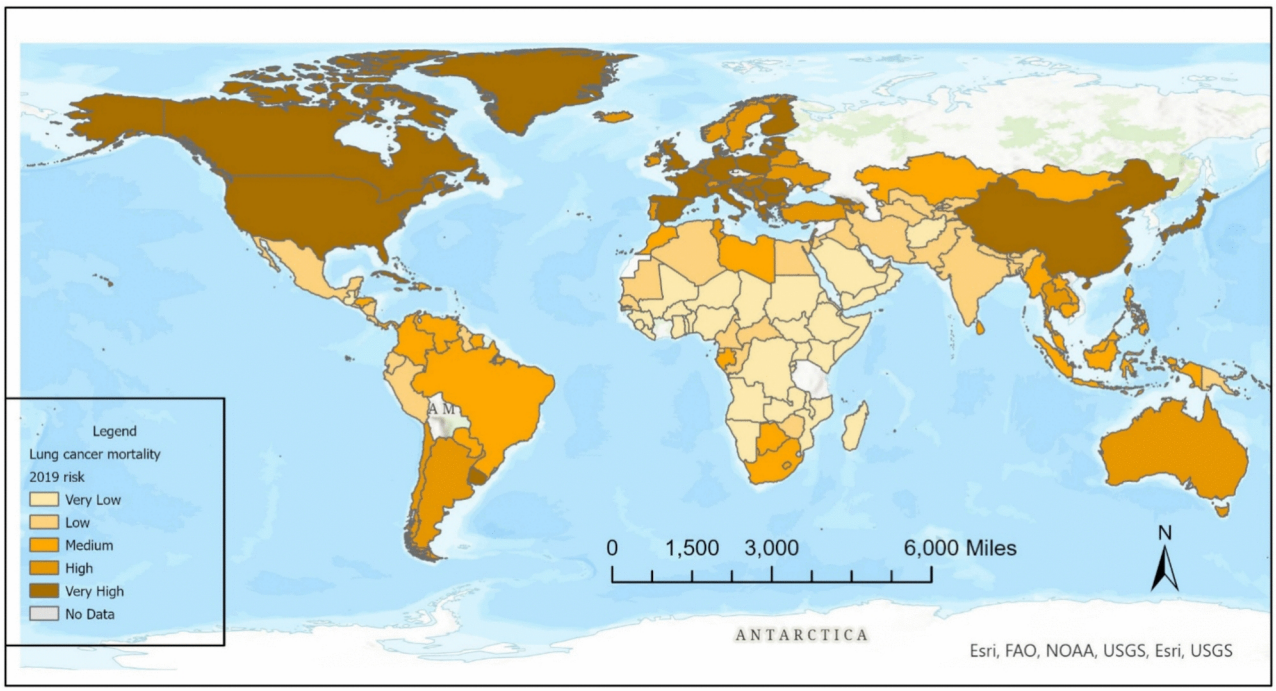

Fig. 1. The distribution of LCM rates across 204 countries.

complex patterns and interaction in the data[17]. Due to flexibility, scalability, and interpretability characteristics, gradient boosting was widely used in epidemiology. The procedure is below.

Format $f_0(\chi)$.

$$f_0(\chi) = arg\min_r \sum_{i=1}^{N} L(y_i, r) \tag{3}$$

Calculate negative gradient for m = 1 to M.

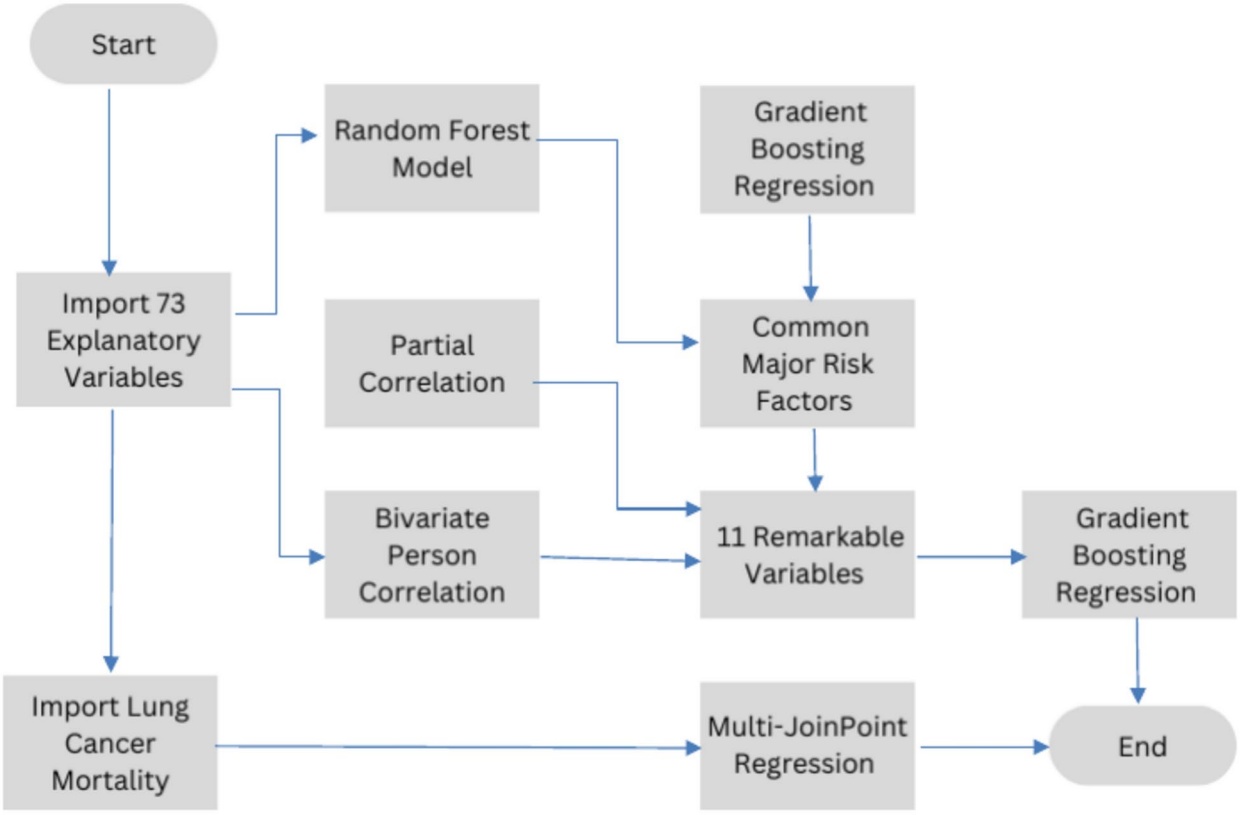

Fig. 2. Analysis framework Note: Produced using Visio Microsoft https://www.microsoft.com/en-us/microsoft-365/visio/flowchart-software (Accessed on October 16th 2024).

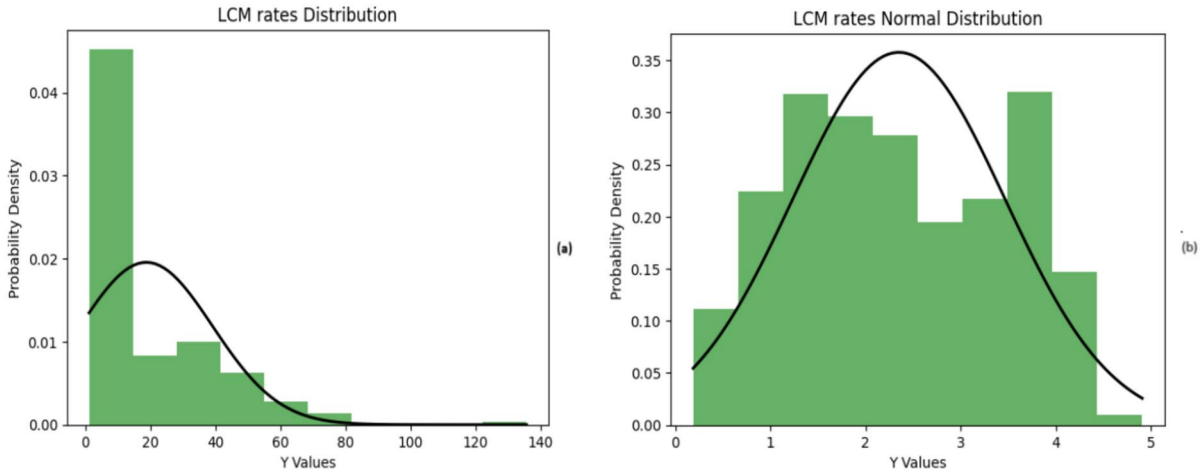

Fig. 3. The comparison of between the LCM rates Distribution and the LCM rates Normal Distribution.

$$y_i\prime = -\frac{\theta L\left(y_i, f_{m-1}\left(x_i\right)\right)}{\theta f_{m-1}\left(x_i\right)} \quad (i = 1, 2, \ldots..., \mathrm{N}) \tag{4}$$

By minimizing Mean Squared Error (MSE), fit $h_m(x)$ into $y_i\prime$.

$$W_m = arg\min_{w} \sum_{i=1}^{N} \left(y_i\prime - h_m\left(x_i, w\right)\right)^2 \tag{5}$$

Use line search to identify learning rate ρ_m to optimize a loss function L.

$$\rho_m = arg\min_{\rho} \sum_{i=1}^{N} L(y_i, f_{m-1}\left(x_i\right) + \rho h_m(x_i, w_m)) \tag{6}$$

$$f_m\left(x\right) = f_{m-1}\left(x\right) + \rho_m h_m(x, w_m) \tag{7}$$

Output $f_m\left(x\right)$.

In Fig. 4, the top five risk factors associated with LCM were occupational exposure to arsenic, smoking, residential radon, occupational exposure to silica and childhood sexual abuse. Of note, after running 73 predisposing factors in the GBR model, occupational exposure to arsenic, smoking, residential radon, occupational exposure to silica four factors were common to the RFT. Therefore, these four factors were treated as control variables in this research.

Joinpoint regression

It is a common modeling in medical research and applied to find the years when mortality rate/incidence case/annual percent change occurred and trend prediction, proposed by Kim (2000)[18–22]. We implemented the Joinpoint regression in Joinpoint Regression Program 4.9.1.0 to estimate trends in LCM rates in a number of countries based on average annual percentage change (AAPC). We used age-adjusted rate of LCM as the

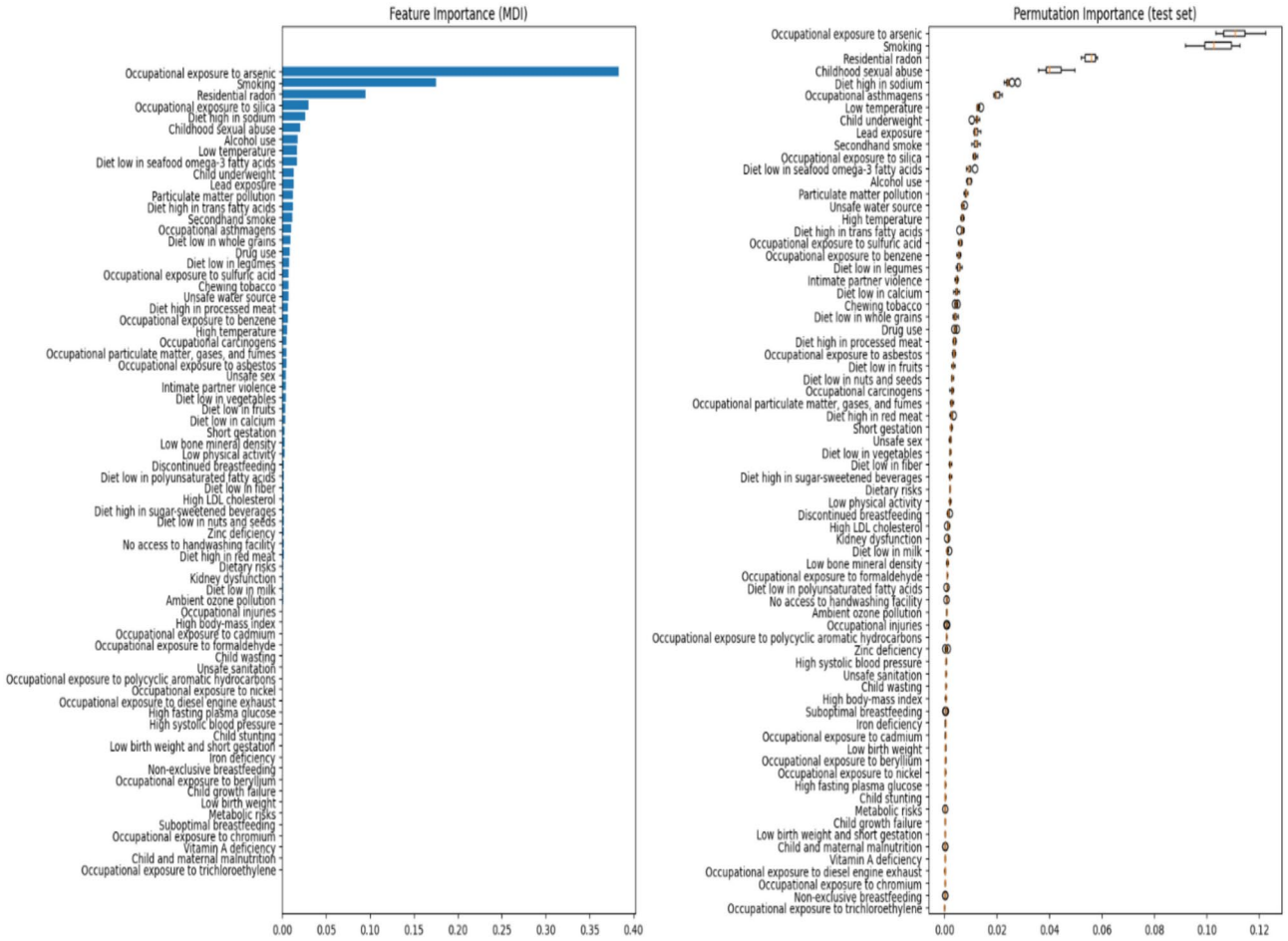

Fig. 4. Importance of different risk factors determined by GBR in 73 Variables.

dependent variable, particulate matter pollution as the standard error, uncorrelated as "choose an error model to fit". The independent variable was "year" and the interval type was "annual". The minimum number of Joinpoints set up as 0 and maximum number was 2(i.e., the maximum number of the infliction point of the trend that can be accepted). The grid search option is selected under the Method settings window. The permutation test was defined with an overall significance level of 0.05 and the default number of permutations set to 4499. Annual Percent Change (APC)/AAPC confidence intervals are selected by parametric method.

Results

Results of RFT

According to the method in the published paper[23], we performed the RFT. The top 5 risk factors associated with LCM were occupational exposure to arsenic, occupational exposure to silica, residential radon, smoking, and occupational exposure to nickel in Fig. 5.

Results of partial correlation

By eliminating the effects of common major factors, 70.59% of the variables' correlation directions were changed. The correlation coefficients for 30 variables, initially positive, became small and negative in the partial correlations, indicating that the impact of these variables may be minimal and not associated with an increase in LCM rates. For example, the Pearson correlation coefficient for ambient ozone pollution, initially 0.342, changed to −0.118 in the partial correlation, suggesting that ambient ozone pollution is not related to an increase in LCM rates. Similarly, some dietary factors, such as a diet low in fruits and a diet low in fiber, also showed exhibited the same pattern.

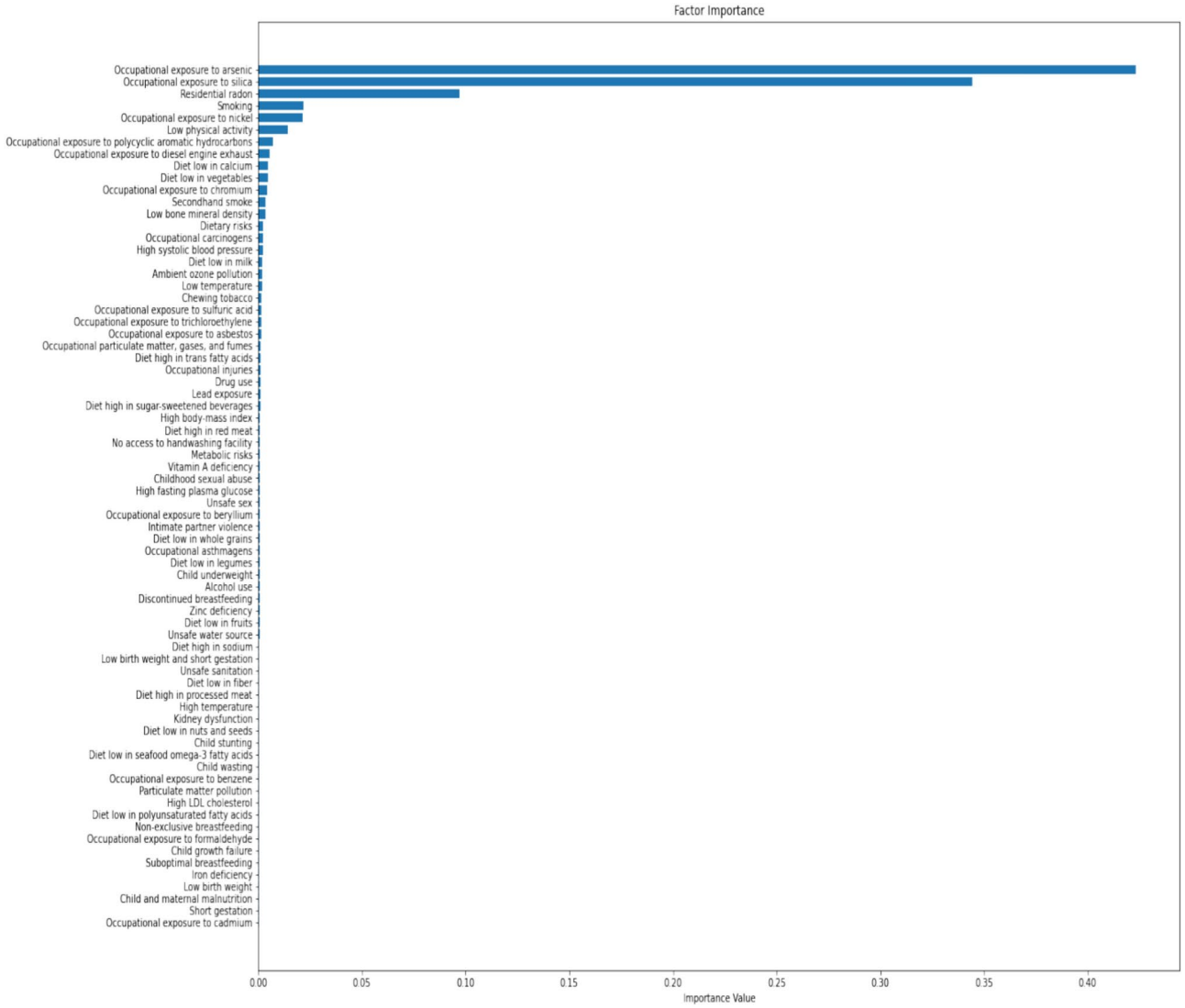

Fig. 5. Importance of different risk factors determined by the FRT. Note: Produced using google Colab (https://colab.research.google.com/, accessed on October 16th 2024).

Additionally, the correlation signs of 17 variables shifted from negative to positive, as shown in Table 2. This change suggests that the original correlations might have been influenced by four major factors. By controlling for these factors, partial correlation reveals the direct relationship between variables. The sign change also indicates a distinction between direct and indirect effects, where the initial correlation may have represented an indirect effect mediated by major factors. Once these major factors are controlled for, the direct effect becomes isolated and may show a different direction. However, their coefficients were lower than 0.1, indicating that their effects were extremely weak and could be disregarded. The coefficients for occupational exposure to asbestos and occupational carcinogens were approximately 0.5, indicating a moderate positive association with LCM.

Results of the selection of variables

We identified eleven significantly representative variables based on prior research findings, ensuring that the selected variables reflect diverse and impactful aspects pertinent to the research objective, as shown in Table 3.

First, alcohol consumption was found to be modestly associated with an increased risk of lung cancer in the pooled analysis. This link is primarily attributed to alcohol's role in increasing oxidative stress or acting as a solvent for tobacco carcinogens[24,25.] Additionally, child stunting and wasting, which result from malnutrition during childhood, have been shown to negatively impact lung development and increase susceptibility to respiratory infections. This contributes to 45% of all deaths in children under five, equating to approximately 3.1 million deaths annually[26,37].

Second, red meat consumption has been significantly associated with a greater risk of lung cancer (RR = 1.26; 95% CI = 1.09–1.44). The risk is linked to high-temperature cooking processes that produce mutagenic heterocyclic amines (HCAs) and polycyclic aromatic hydrocarbons (PAHs), and the heme iron content in red meat, which can catalyze the formation of N-nitroso compounds (NOCs)[27]. In contrast, a higher body mass index (BMI) appears to correlate with a reduced risk of lung cancer, likely due to confounding factors such as smoking-related weight loss and reverse causation, where preclinical lung cancer leads to lower body weight[28].

Third, elevated low-density lipoprotein (LDL) cholesterol has been implicated in lung cancer development. LDL can enhance cancer cell growth and metastasis by modulating membrane fluidity and signaling pathways[29]. Both systolic and diastolic blood pressures were identified as significant predictors of lung cancer, with a 10% increase in risk for every 10-mmHg increment in blood pressure. This is due to shared risk factors such as smoking and underlying metabolic abnormalities[30].

Fourth, metabolic risks, including obesity, hypertension, and insulin resistance, have been linked to an increased risk of lung cancer[31]. Occupational exposure to asbestos is another significant factor, with studies indicating that asbestos-related lung cancer (ARLC) incidence is six times larger than mesothelioma incidence, marking it as a critical health concern[32].

Fifth, particulate matter (PM) pollution, especially fine particles like PM2.5, has been established as a significant risk factor for lung cancer. The International Agency for Research on Cancer (IARC) has classified outdoor air pollution and PM2.5 as carcinogenic to humans, affirming the clear link between PM exposure and lung cancer development[33]. Furthermore, a systematic review published in the *European Journal of Public Health*estimated that ambient PM2.5 exposure contributes to a substantial burden of lung cancer cases globally, emphasizing the critical need for air quality interventions to reduce cancer risk[34].

Lastly, secondhand smoke, a mixture of side stream and mainstream tobacco smoke, contains over 7,000 chemicals, including around 70 known carcinogens. Inhalation of this smoke increases the risk of lung cancer in nonsmokers by 20–30%, leading to over 7,300 lung cancer deaths annually among U.S. adults who do not smoke. The correlation between secondhand smoke and lung cancer is due to the toxic and carcinogenic substances present in the smoke, such as benzene, formaldehyde, and polycyclic aromatic hydrocarbons. These findings are based on the National Academy Press (1986), which provided crucial insights into the harmful effects of secondhand smoke[35].

Results of gradient boosting regression

Through Gradient Boosting Regressor with least squares loss and 500 regression trees of depth 4 in supplemental Fig. 1, the mean squared error (MSE) was 9.36 and R^2 was 0.98. We set parameters including n_estimators of 500 (i.e., boosting stages were 500.), max_depth of 4 (i.e., the number of nodes in the tree is limited to 4.), min_samples_split of 5 (i.e., the minimum number of samples required to split an internal node was 5.), learning_rate of 0.01 (i.e., the contribution of each tree would shrink by 0.01.), loss of squared_error (i.e., the least squares function was used as the loss function to optimize.). In supplemental Fig. 2, deviance in training set and test set were close to 0 and dramatically decreased, which training set deviance was highly fit in test set deviance. The reason was that deviances, as the opposite of variance, were less and less, which meant the predicted regression was better and better. Supplemental Fig. 3 displayed permutation importance of 11 extracted risk factors determined by GBR. The significance of occupational exposure to asbestos, high systolic blood pressure, secondhand smoke, child wasting, ,alcohol use, child stunting, diet high in red meat, and high LDL cholesterol was found to be greater than that of particulate matter pollution (PM2.5).

Results of joinpoint regression

A multiple join-point regression analysis of LCM was conducted in 204 countries in Fig. 6, we merely presented a part of countries such as Australia, Brazil, Canda, China, Cuba, Denmark, Germany, and Japan, owing to the shared characteristics of trend homogeneity. During the study period, Canada, France, Japan, the U.S and etc. have a single infliction point, while China, Germany, Indonesia, Norway and etc. have two the infliction points in Table 4. Denmark ranked first in LCM rates while Sudan ranked last. LCM in China, Indonesia, Bermuda, and Japan was considered to have a significantly increasing trend due to the positive APC, whereas the LCM in Denmark was considered to have a significantly decreasing trend due to a negative APC in Table 5. Regarding

Factor	Pearson Correlation (Coe./ Sig.)	partial Correlation (Coe./ Sig.)
Ambient ozone pollution	0.342/0.000	−0.118/0.000
Alcohol use	0.505/0.000	−0.213/0.000
Chewing tobacco	−0.114/0.000	−0.110/0.000
Child and maternal malnutrition	−0.405/0.000	0.036/0.046
Child growth failure	−0.389/0.000	0.046/0.011
Child stunting	−0.354/0.000	0.049/0.006
Child underweight	−0.367/0.000	0.034/0.059*
Child wasting	−0.389/0.000	0.046/0.011
Childhood sexual abuse	0.326/0.000	−0.169/0.000
Diet high in processed meat	0.644/0.000	−0.037/0.04
Diet high in red meat	0.696/0.000	−0.137/0.000
Diet high in sodium	0.497/0.000	−0.28/0.000
Diet high in sugar-sweetened beverages	0.536/0.000	−0.108/0.000
Diet high in trans fatty acids	0.513/0.000	−0.229/0.000
Diet low in calcium	0.455/0.000	−0.236/0.000
Diet low in fiber	0.346/0.000	−0.342/0.000
Diet low in fruits	0.255/0.000	−0.373/0.000
Diet low in legumes	0.456/0.000	−0.248/0.000
Diet low in milk	0.658/0.000	−0.102/0.000
Diet low in nuts and seeds	0.336/0.000	−0.221/0.000
Diet low in polyunsaturated fatty acids	0.300/0.000	−0.279/0.000
Diet low in seafood omega-3 fatty acids	0.440/0.000	−0.316/0.000
Diet low in vegetables	0.031/0.084*	−0.264/0.000
Diet low in whole grains	0.513/0.000	−0.182/0.000
Dietary risks	0.576/0.000	−0.365/0.000
Discontinued breastfeeding	−0.349/0.000	0.058/0.001
Drug use	0.352/0.000	−0.098/0.000
High body-mass index	0.467/0.000	−0.181/0.000
High fasting plasma glucose	0.401/0.000	−0.134/0.000
High LDL cholesterol	0.568/0.000	−0.306/0.000
High systolic blood pressure	0.568/0.000	−0.201/0.000
High temperature	−0.419/0.000	−0.018/0.320*
Intimate partner violence	−0.208/0.000	0.064/0.000
Iron deficiency	−0.449/0.000	0.038/0.034
Kidney dysfunction	0.442/0.000	−0.162/0.000
Lead exposure	0.105/0.000	−0.113/0.000
Low birth weight	−0.526/0.000	0.030/0.096
Low birth weight and short gestation	−0.525/0.000	0.032/0.073*
Low bone mineral density	0.665/0.000	−0.096/0.000
Low temperature	0.728/0.000	−0.099/0.000
Metabolic risks	0.597/0.000	−0.262/0.000
No access to handwashing facility	−0.413/0.000	0.073/0.000
Non-exclusive breastfeeding	−0.366/0.000	0.053/0.003
Occupational asthmagens	−0.445/0.000	−0.125/0.000
Occupational carcinogens	0.837/0.000	0.480/0.000
Occupational exposure to asbestos	0.792/0.000	0.478/0.000
Occupational exposure to benzene	0.236/0.000	−0.068/0.000
Occupational exposure to beryllium	0.396/0.000	0.003/0.883*
Occupational exposure to cadmium	0.464/0.000	−0.111/0.000
Occupational exposure to chromium	0.473/0.000	−0.057/0.002
Occupational exposure to diesel engine exhaust	0.385/0.000	0.059/0.001
Occupational exposure to formaldehyde	−0.215/0.000	−0.228/0.000
Occupational exposure to polycyclic aromatic hydrocarbons	0.470/0.000	−0.022/0.231*
Occupational exposure to sulfuric acid	0.499/0.000	−0.055/0.002
Occupational exposure to trichloroethylene	0.264/0.000	−0.130/0.000
Continued		

Factor	Pearson Correlation (Coe./ Sig.)	partial Correlation (Coe./ Sig.)
Occupational injuries	−0.357/0.000	−0.123/0.000
Occupational particulate matter, gases, and fumes	0.237/0.000	−0.25/0.000
Particulate matter pollution	−0.185/0.000	−0.115/0.000
Secondhand smoke	0.370/0.000	−0.168/0.000
Short gestation	−0.528/0.000	0.039/0.031
Suboptimal breastfeeding	−0.366/0.000	0.053/0.003
Unsafe sanitation	−0.385/0.000	0.048/0.008
Unsafe sex	−0.220/0.000	0.059/0.001
Unsafe water source	−0.396/0.000	0.046/0.012
Vitamin A deficiency	−0.263/0.000	0.043/0.018
Zinc deficiency	−0.272/0.000	0.019/0.301*
Household air pollution	−0.049/0.007	−0.065/0.000

Table 2. The Comparison Table between Partial and Bivariate Correlation. Note: 1. * donates not significant.

No	Factor	Pearson Correlation (Coe. / Sig.)	Partial Correlation (Coe. / Sig.)
1	Alcohol use	0.505/0.000	−0.213/0.000
2	Child stunting	−0.354/0.000	0.049/0.006
3	Child wasting	−0.389/0.000	0.046/0.011
4	Diet high in red meat	0.696/0.000	−0.137/0.000
5	High body-mass index	0.467/0.000	−0.181/0.000
6	High LDL cholesterol	0.568/0.000	−0.306/0.000
7	High systolic blood pressure	0.568/0.000	−0.201/0.000
8	Metabolic risks	0.597/0.000	−0.262/0.000
9	Occupational exposure to asbestos	0.792/0.000	0.478/0.000
10	Particulate matter pollution	−0.185/0.000	−0.115/0.000
11	Secondhand smoke	0.370/0.000	−0.168/0.000

Table 3. Significantly Impactful Variables List.

the Average Annual Percent Change (AAPC) in LCM rates, significant increasing trends were observed in 142 countries, including Bermuda, China, France, Indonesia, and Japan, as indicated by the positive AAPC values shown in Table 6. Conversely, LCM rates in 38 countries, such as Norway and Demark exhibited a decreasing trend, while 24 countries, including Sudan, Mali, and Australia demonstrated stable trends over the same period, as also outlined in Table 6.

Discussion

The analysis identified four major factors contributing to lung cancer: occupational exposure to arsenic, smoking, residential radon exposure, and occupational exposure to silica. Additionally, the significance of occupational exposure to asbestos, high systolic blood pressure, secondhand smoke, child wasting, and alcohol use was found to be greater than that of particulate matter pollution (PM2.5). It does not mean PM2.5 less important. PM2.5, meaning 2.5 μm or smaller in diameter, has significant impacts on lung cancer due to its ability to penetrate deep into the lungs and even enter the bloodstream. PM2.5 was reported that it had high burden of acute respiratory infections (ARI) in children[36]. Among the countries studied, China, France, Indonesia, and Japan showed increasing trends in LCM rates, while Norway demonstrated a decreasing trend. These findings enhance our understanding of early prevention and holistic recognition of lung cancer at the country level. The insights provided by the proposed tree-based machine learning algorithm, in conjunction with partial correlation and multiple Joinpoint regression analysis, can be applied to other significant diseases to evaluate multifaceted factors that are closely related. Meanwhile, the research provided additional perspective into the GBD risk factors for interpreting fatal diseases of LCM rates. GBD is a global project involving more than 5,000 researchers from 154 countries and territories on 483 separate outcomes of 107 diseases and injuries and 14 million death certificates[37–40].

First, The identification of occupational exposure to arsenic, smoking, residential radon, and occupational exposure to silica as four major common factors influencing LCM rates corroborates findings from prior lung cancer research. Generally speaking, arsenic, radon, and silica as toxic chemical elements do not directly affect humans but rather infiltrate and compromise health through the surrounding environment. They permeate outdoor air, indoor environments, everyday materials, drinking water, and other mediums, subtly encroaching on people's well-being. Arsenic, is a carcinogenic chemical, involving long latency periods and indirect effects on

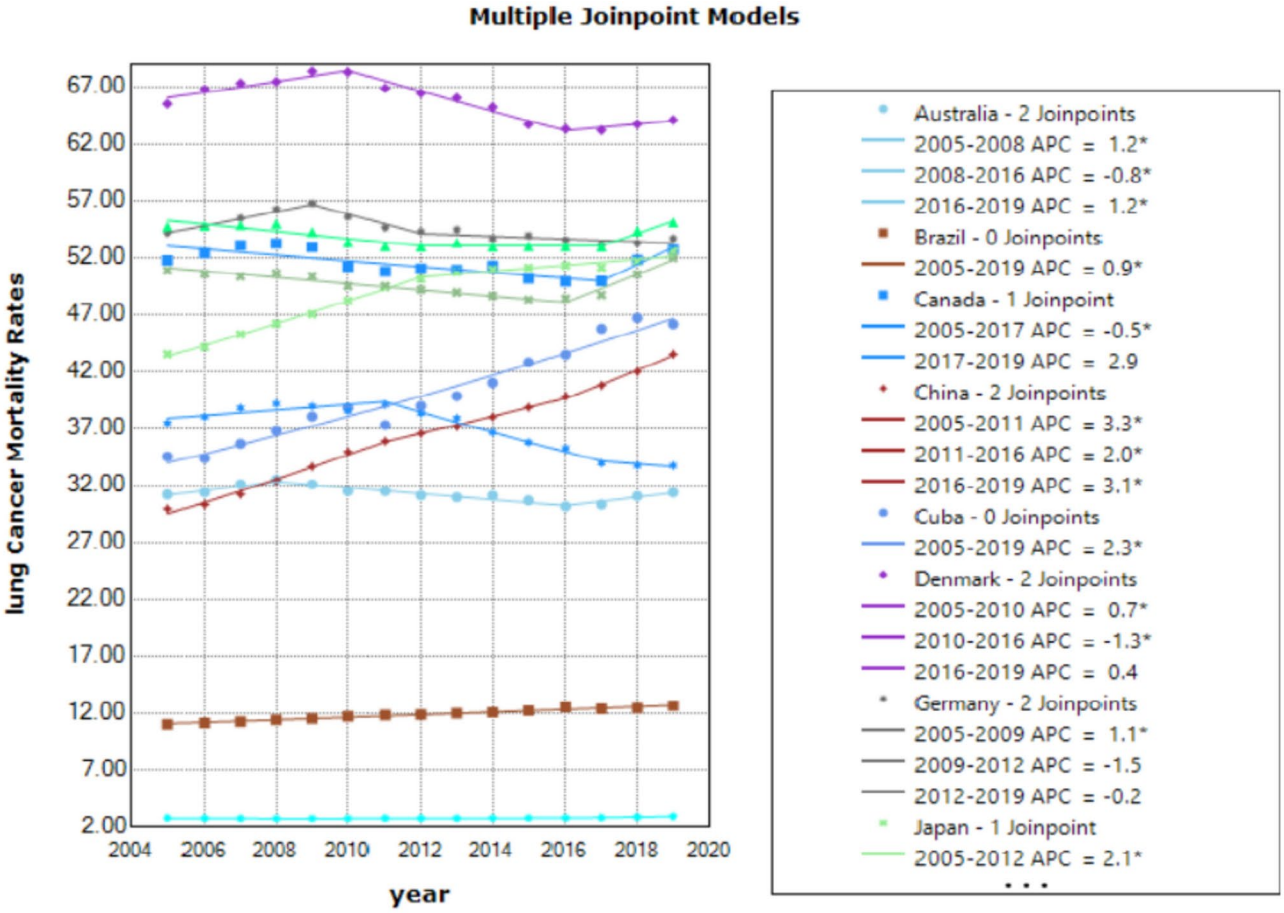

Fig. 6. A Multiple Joinpoint Regression Model.

Cohort	Joinpoint	Estimate	Lower CI	Upper CI
Canada—1 Joinpoint	1	2017	2015	2017
China—2 Joinpoints	1	2011	2007	2013
China—2 Joinpoints	2	2016	2010	2017
France—1 Joinpoint	1	2012	2011	2014
Germany—2 Joinpoints	1	2009	2007	2010
Germany—2 Joinpoints	2	2012	2010	2017
Indonesia—2 Joinpoints	1	2011	2010	2013
Indonesia—2 Joinpoints	2	2015	2014	2017
Japan—1 Joinpoint	1	2012	2011	2013
Norway—2 Joinpoints	1	2011	2007	2013
Norway—2 Joinpoints	2	2017	2010	2017
United States of America—1 Joinpoint	1	2016	2015	2017

Table 4. Estimated Joinpoints.

human health. It typically through inhalation in industrial settings such as mining, smelting, and manufacturing, can lead to the toxic accumulation of arsenic in the lungs. Prolonged exposure can damage DNA and induce oxidative stress, ultimately increasing the risk of developing lung cancer. Irva et al. (1992) mentioned that the joint effects of both occupational exposure to arsenic and smoking increased the risk of lung cancer from 70 to 130%[41]. Radon is a radioactive gas that naturally arises from the decay of uranium found in soil, rock, and water. It can build up inside buildings, especially in basements and ground floors. Inhaling radon decay products can harm lung tissue, causing mutations and raising the risk of lung cancer. Radon concentrations also caused a significant risk of lung cancer[8]. Silica dust is produced in industries like construction, mining, and manufacturing. Breathing in fine silica particles can lead to silicosis, a lung disease characterized by scarring of

Cohort	Segment	Lower Endpoint	Upper Endpoint	APC	CI	t Test	Prob>|t|
Canada-1 Joinpoint	1	2005	2017	−0.495*	[−0.741, −0.249]	−4.48	0.001
Canada-1 Joinpoint	2	2017	2019	2.928	[−0.524, 6.500]	1.885	0.089
China-2 Joinponts	1	2005	2011	3.264*	[2.816, 3.715]	17.451	< 0.001
China-2 Joinponts	2	2011	2016	2.047*	[1.357, 2.741]	7.066	< 0.001
China-2 Joinponts	3	2016	2019	3.075*	[2.119, 4.040]	7.682	< 0.001
France-1 Joinpoint	1	2005	2012	2.068*	[1.599, 2.538]	9.921	< 0.001
France-1 Joinpoint	2	2012	2019	−0.509*	[−0.849, 0.167]	−3.317	0.008
Germany-2 Joinponts	1	2005	2009	1.131*	[0.407, 1.861]	3.701	0.008
Germany-2 Joinponts	2	2009	2012	−1.531	[−3.812, 0.804]	−1.557	0.163
Germany-2 Joinponts	3	2012	2019	−0.219	[−0.472, 0.034]	−2.044	0.08
Indonesia-2 Joinponts	1	2005	2011	2.808*	[2.702, 2.914]	63.688	< 0.001
Indonesia-2 Joinponts	2	2011	2015	2.218*	[1.952, 2.484]	19.943	< 0.001
Indonesia-2 Joinponts	3	2015	2019	2.973*	[2.833, 3.114]	50.91	< 0.001
Japan-1 Joinpoint	1	2005	2012	2.14*	[1.953, 2.327]	25.778	< 0.001
Japan-1 Joinpoint	2	2012	2019	0.509*	[0.305, 0.713]	5.562	< 0.001
Norway-2 Joinponts	1	2005	2011	0.627	[−0.038, 1.296]	2.229	0.061
Norway-2 Joinponts	2	2011	2017	−2.344*	[−2.938, −1.746]	−9.193	< 0.001
Norway-2 Joinponts	3	2017	2019	−0.708	[−2.932, 1.567]	−0.741	0.483
The U.S.−1 Joinpoint	1	2005	2016	−0.551*	[−0.645, −0.457]	−13.079	< 0.001
The U.S.−1 Joinpoint	2	2016	2019	2.508*	[2.007, 3.012]	11.254	< 0.001

Table5. Annual Percent Change List. *Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.

Cohort	Lower Endpoint	Upper Endpoint	AAPC	Lower CI	Upper CI	Test Statistic	P-Value
Canada—1 Joinpoint	2005	2019	−0.013	−0.479	0.455	−0.055	0.956
China—2 Joinpoints	2005	2019	2.787*	2.477	3.099	17.804	< 0.001
France—1 Joinpoint	2005	2019	0.771*	0.517	1.025	5.97	< 0.001
Germany—2 Joinpoints	2005	2019	−0.119	−0.579	0.343	−0.505	0.614
Indonesia—2 Joinpoints	2005	2019	2.686*	2.606	2.767	66.159	< 0.001
Japan—1 Joinpoint	2005	2019	1.321*	1.199	1.443	21.384	< 0.001
Norway—2 Joinpoints	2005	2019	−0.847*	−1.26	−0.432	−3.991	< 0.001
United States of America—1 Joinpoint	2005	2019	0.097	−0.016	0.21	1.677	0.093

Table 6. Average Annual Percent Change List (AAPC). *Indicates that the AAPC is significantly different from zero at the alpha = 0.05 level.

lung tissue. This scarring creates conditions that promote the development of lung cancer by inducing chronic inflammation and increasing cellular damage. The relationship between occupational exposure to silica and lung cancer risk had a positive exposure–response relationship[42]. Of note, the identification of smoking as one of the main risk factors highlights the credibility of our methodology, as it aligns with the findings of most existing lung cancer studies.

Second, our analysis revealed that poccupational exposure to asbestos, high systolic blood pressure, secondhand smoke, child wasting, and alcohol use were the top five variables that had a greater impact on LCM rates than environmental factors, with the exception of the four common risk factors, among the 73 factors evaluated using Gradient Boosting Regression. Some literatures mentioned occupational exposure to asbestos, High systolic blood pressure, Secondhand smoke, Child wasting, and Alcohol use exacerbated the risk of lung cancer[43–47]. Asbestos refers to fibrous minerals, which used to resist either heat or corrosion. When asbestos fibers are absorbed, they can be stored in the lungs, resulting in inflammation and scarring to form asbestosis, which is a chronic lung disease, increasing the risk of lung cancer. Asbestosis in asbestos-exposed lung cancer account for more than in the nonexposed lung cancer[7]. High systolic blood pressure is generally associated with cardiovascular diseases, recent studies suggest a potential link to lung cancer. Cho et al. (2021) mentioned high systolic blood pressure was linked increased risk of lung cancer[48]. Secondhand smoke contains benzene, formaldehyde, and nitrosamines carcinogens. Regular exposure for non-smokers to secondhand smoke can do harmful lung tissue, causing mutations that enhance the risk of lung cancer. Some research mentioned secondhand smoke had the most detrimental effects for children[49,50]. Child wasting revealed the importance of nutrition for LCM, which was confirmed by Ilaria et al. (2020)[51]. Alcohol use does attack and weaken individual's immune system, making the body less capable of fighting off cancerous alteration in the lungs and other organs.

Some existed research also mentioned alcohol consumptions was pertinent to lung cancer onset in never-smokers[52–54].

Next, Multiple Joinpoint regression not only summarized the year-to-year variations in LCM rates with different trends across all countries but also provided a valuable tool for comparing disparities between these units during the study period[19]. To maximize the utility of the investigation, we selected 204 countries for analysis. The findings indicate that LCM rates showed increasing trends in 142 countries. These countries are often highly populated, such as China, India, and Indonesia, and span various climates-from tropical regions like Indonesia and Brazil to temperate zones like Finland and Sweden, and even desert areas such as Saudi Arabia and Namibia. These regions often face health challenges due to limited distribution of health resources. Those countries should pay more attention to improve health system and LCM rates prevention. Conversely, LCM rates exhibited decreasing trends in 38 countries, which are typically developed nations facing different public health challenges. Meanwhile, 24 countries showed stable LCM trends, suggesting that nations with more advanced healthcare systems tend to achieve better health outcomes, whereas those with weaker systems may struggle with higher LCM rates. Additionally, this study focused on analyzing longitudinal trends in LCM rates for vertical comparisons and examined the potential impact of predisposing factors through horizontal contrasts. This approach aims to enhance the understanding of early prevention and control of lung cancer at the national level.

Lastly, our analysis has several limitations. Firstly, the GBD data were large scale, aggregated at the country level, encompassing a vast array of exposure estimates that caused the inconsistency between localized findings and global assessments should be addressed in future research. For instance, Smith et al. (2018) reported an increase in LCM associated with arsenic exposure (relative risk [RR] of 3.38) in regions of Chile, focusing on the long-term impact of arsenic exposure on LCM using limited datasets[55]. Nevertheless, Leiter et al. (2023) identified smoking as the leading risk factor for lung cancer[56,57]. In our analysis, the risk associated with smoking appears lower than that of arsenic exposure. This discrepancy may stem, in part, from differences in study scale, data sources, and the limitations of machine learning methods, which may not fully account for discrepancies in measurement units. Addressing this discrepancy, we should further detail how occupational arsenic exposure was measured and whether the statistical model adjusted appropriately for confounders. Moreover, LCM research should further consider demographic factors such as gender, age, and racial differences across different countries. However, current lung cancer research predominantly focuses on macro-level analyses of country heterogeneity, rather than addressing and targeting the core aspects of lung cancer therapy. Furthermore, LCM should be categorized into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) for a more nuanced discussion, given that NSCLC and SCLC account for 85% and 15% of lung cancer cases, respectively[14,58,59]. Finally, this study may have a temporal disconnect between the exposure to certain risk factors and the observed lung cancer outcomes. Specifically, the latency period for asbestos-related lung cancer is typically 20–40 years. This means that exposure to asbestos in earlier decades, such as the 1960s-1980s, would most likely be responsible for the LCM observed in the 2004–2020 period. Therefore, it is important to recognize that the lung cancer outcomes in this study reflect past exposures rather than current or recent risk factor levels.

Conclusion

Lung cancer is an enteral academic spotlight, unless cancer is no longer a threat to humanity. How to mitigate LCM is still a challenge. Our study utilized machine learning algorithms to identify four common major risk factors (i.e., occupational exposure to arsenic, smoking, residential radon, occupational exposure to silica) and high impactful risk factors (i.e., Occupational exposure to asbestos, High systolic blood pressure, Secondhand smoke, Child wasting, and Alcohol use), as well as Multiple Joinpoint Regression to identify increasing trends of LCM rates in 142 countries (e.g., China and India); decreasing trends in 38 countries (e.g., Denmark and Norway), and stable trends in 24 countries (e.g., Sudan, Mali, and Australia). This research suggests that in addition to considering the effects of occupational exposure to arsenic, smoking, residential radon, and occupational exposure to silica on LCM rates, occupational exposure to asbestos, high systolic blood pressure, secondhand smoke, child wasting, and alcohol use should be considered in lung cancer prevention strategies in countries with increasing trends of LCM rates. Future research should focus on determining the impact of these risk factors on LCM rates at mesoscale and fine scale.

Data availability

All data, models, and code generated or used during the study appear in the submitted article.

Received: 9 June 2024; Accepted: 27 February 2025
Published online: 13 March 2025

References

1. Choi, E. et al. Risk model-based management for second primary lung cancer among lung cancer survivors through a validated risk prediction model. *Cancer* **130**(5), 770–780. https://doi.org/10.1002/cncr.35069 (2024).
2. Molassiotis, A., Kwok, S. W. H., Leung, A. Y. M. & Tyrovolas, S. Associations between sociodemographic factors, health spending, disease burden, and life expectancy of older adults (70 + years old) in 22 countries in the Western Pacific Region, 1995–2019: estimates from the Global Burden of Disease (GBD) Study 2019. *GeroScience* **44**(2), 925–951. https://doi.org/10.1007/s11357-021-00494-z (2022).
3. Siegel, R. L., Miller, K. D. & Fuchs, H. E. Cancer statistics. *CA: A Cancer J. Clin.* **72**(1), 7–33. https://doi.org/10.3322/caac.217 (2022).
4. Ju, W. et al. Cancer statistics in Chinese older people : current burden, time trends, and comparisons with the US, Japan, and the Republic of Korea. *Sci. China Life Sci.* https://doi.org/10.1007/s11427-022-2218-x (2022).
5. Samet, J. M. Carcinogenesis and lung cancer: 70 years of progress and more to come. *Carcinogenesis.* **41**(10), 1309–1317. https://doi.org/10.1093/carcin/bgaa094 (2020).

6. Omenn, G. S. et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N. Engl. J. Med.* **334**(18), 1150–1155 (1996).
7. Icksan, A. G., Widharisastra, C., Suraya, A., & Ferstl, M. Pleural plaques and pleural changes among lung cancer patients exposed to asbestos. Universa Medicina 41 3 210–218 https://doi-org.libproxy.txstate.edu/https://univmed.org/ejurnal/index.php/medicina/article/view/1307/712 (2022)
8. Alavanja, M. C. R., Lubin, J. H., Mahaffey, J. A. & Brownson, R. C. Residential radon exposure and risk of lung cancer in missouri. *Am. J. Public Health* **89**(7), 1042–1048. https://doi.org/10.2105/AJPH.89.7.1042 (1999).
9. Jones, R. N., Hughes, J. M. & Weill, H. Asbestos exposure, asbestosis, and asbestos-attributable lung cancer. *Thorax* **51**(suppl 2), S9–S12 (1996).
10. Luo, R. X., Wu, B., Yi, Y. N., Huang, Z. W. & Lin, R. T. Indoor burning coal air pollution and lung cancer—a case-control study in Fuzhou. *China. Lung Cancer* **14**(suppl 1), S113–S119 (1996).
11. Elizabeth, G. New evidence that air pollution contributes substantially to lung cancer. *The Lancet Oncol.* **23**(10), e448. https://doi.org/10.1016/S1470-2045(22)00569-1 (2022).
12. Pope, C. A. et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc.* **287**(9), 1132–1141 (2002).
13. Bray, F. I. & Weiderpass, E. Lung cancer mortality trends in 36 European countries: secular trends and birth cohort patterns by sex and region 1970–2007. *Int. J. Cancer* **126**(6), 1454–1466. https://doi.org/10.1002/ijc.24855 (2010).
14. Gongbo, C., Xia, W., Gonghuan, Y. & Xiaonong, Z. Traffic-related air pollution and lung cancer: a meta-analysis. *Thorac. Cancer* **6**, 307–318 (2015).
15. Zeng, X., Zhu, S., Xu, C., Wang, Z., Su, X., Zeng, D., Long, H., & Zhu, B. (2020). Effect of Comorbidity on Outcomes of Patients with Advanced Non-Small Cell Lung Cancer Undergoing Anti-PD1 Immunotherapy Medical Science Monitor https://doi.org/10.12659/MSM.922576.
16. Jinting Zhang, F., Zhan, B., Xiu, Wu. & Zhang, D. Partial correlation analysis of association between subjective well-being and ecological footprint. *Sustainability.* **13**(3), 1033. https://doi.org/10.3390/su13031033 (2021).
17. Luo, J., Wei, Z., Man, J. & Xu, S. TRBoost: a generic gradient boosting machine based on trust-region method. *Appl. Intell.: Int. J. Res. Intell. Syst. Real Life Complex Probl.* https://doi.org/10.1007/s10489-023-05000-w (2023).
18. Dragomirescu, I. et al. A join point regression analysis of trends in mortality due to osteoporosis in Spain. *Sci. Rep.* **9**, 4264. https://doi.org/10.1038/s41598-019-40806-0 (2019).
19. Akinyede, O. & Soyemi, K. Joinpoint regression analysis of pertussis crude incidence rates, Illinois, 1990–2014. *Am. J. infect. control.* **44**(12), 1732–1733. https://doi.org/10.1016/j.ajic.2016.05.013 (2016).
20. Kim, H. J., Fay, M. P., Feuer, E. J. & Midthune, D. N. Permutation tests for joinpoint regression with applications to cancer rates. *Stat. Med.* **19**(3), 335–351 (2000).
21. Huang, J. et al. Incidence and mortality of thyroid cancer in 50 countries: A joinpoint regression analysis of global trends. *Endocrine: : Int. J. Basic & Clin. Endocrinol.* https://doi.org/10.1007/s12020-022-03274-7 (2023).
22. Politis, M. et al. Trend analysis of cancer mortality and incidence in Panama Using Joinpoint Regression Analysis. *Medicine* **94**(24), e970. https://doi.org/10.1097/MD.0000000000000970 (2015).
23. Wu, X., Denise, B.-B., Zhan, F. B. & Zhang, J. Determining association between lung cancer mortality worldwide and risk factors using fuzzy inference modeling and random forest modeling. *Int. J. Environ. Res. & public health.* **19**, 21. https://doi.org/10.3390/ijerph192114161 (2022).
24. Freudenheim, J. L. et al. Alcohol consumption and risk of lung cancer: a pooled analysis of cohort studies. *Am. J. Clin. Nutr.* **82**(3), 657–667. https://doi.org/10.1093/ajcn/82.3.657 (2005).
25. Bagnardi, V. et al. Alcohol consumption and lung cancer risk in never smokers: a meta-analysis. *Ann. Oncol.* **22**(12), 2631–2639. https://doi.org/10.1093/annonc/mdr027 (2011).
26. Tharumakunarajah, R., Lee, A., Hawcutt, D. B., Harman, N. L. & Sinha, I. P. The impact of malnutrition on the developing lung and long-term lung health: A narrative review of global literature. *Pulm. Ther.* **10**(2), 155–170. https://doi.org/10.1007/s41030-024-00257-z (2024).
27. Farvid, M. S. et al. Consumption of red meat and processed meat and cancer incidence: a systematic review and meta-analysis of prospective studies. *Eur. J. Epidemiol.* **36**(9), 937–951. https://doi.org/10.1007/s10654-021-00741-9 (2021).
28. Robien, K. & DiPietro, L. Obesity, metabolic syndrome, and lung cancer prevention. *Chest* **165**(1), 14–15. https://doi.org/10.1016/j.chest.2023.10.044 (2024).
29. Kuzu, O. F., Noory, M. A. & Robertson, G. P. The Role of cholesterol in cancer. *Cancer Res. (Chicago, Ill.)* **76**(8), 2063–2070. https://doi.org/10.1158/0008-5472.CAN-15-2613 (2016).
30. Lindgren, A., Pukkala, E., Nissinen, A. & Tuomilehto, J. Blood pressure, smoking, and the incidence of lung cancer in hypertensive men in North Karelia Finland. *Am. J. Epidemiol.* **158**(5), 442–447. https://doi.org/10.1093/aje/kwg179 (2003).
31. Lee, Y.-Y. et al. Association between risk factors of metabolic syndrome with lung function. *Eur. J. Clin. Nutr.* **74**(5), 811–817. https://doi.org/10.1038/s41430-018-0369-6 (2020).
32. van Zandwijk, N., Frank, A. L., Reid, G., Dimitri Røe, O. & Amos, C. I. Asbestos-related lung cancer: An underappreciated oncological issue. *Lung Cancer (Amsterdam, Netherlands)* **194**, 107861. https://doi.org/10.1016/j.lungcan.2024.107861 (2024).
33. Berg, C. D. et al. Air pollution and lung cancer: A review by international association for the study of lung cancer early detection and screening committee. *J. Thorac. Oncol.* **18**(10), 1277–1289. https://doi.org/10.1016/j.jtho.2023.05.024 (2023).
34. Corda, M. et al. The burden of lung cancer attributable to ambient PM2.5 exposure: A systematic literature review. *Eur. J. Public Health* https://doi.org/10.1093/eurpub/ckae144.1418zs (2024).
35. *Environmental tobacco smoke : Measuring exposures and assessing health effects* (1st ed.) (National Academy Press, 1986)
36. Adhikary, M., Mal, P. & Saikia, N. Exploring the link between particulate matter pollution and acute respiratory infection risk in children using generalized estimating equations analysis: a robust statistical approach. *Environ. Health: A Global Access Sci. Source* **23**(1), 1–11. https://doi.org/10.1186/s12940-024-01049-3 (2024).
37. Murray, C. J. L. et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *The Lancet* **396**(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2 (2020).
38. Solmi, M. et al. Incidence, prevalence, and global burden of schizophrenia - data, with critical appraisal, from the Global Burden of Disease (GBD) 2019. *Mol. Psychiatr.* **28**(12), 5319–5327 (2023).
39. Ortblad, K. F., Lozano, R. & Murray, C. J. The burden of HIV: insights from the Global Burden of Disease (GBD) study 2010. *The Lancet (British Edition)* **381**(S2), S103–S103. https://doi.org/10.1016/S0140-6736(13)61357-4 (2013).
40. Nichols, E., Wang, C. & Vos, T. P1–540: Multidimensional Item Response Theory (IRT) For survey harmonization and dementia prevalence prediction in the Global Burden of Disease (GBD) Study: An example using the Health And Retirement An Retirement Survey (HRS). *Alzheimer's & Dement.* **15**(7S_Part_9), P478–P478. https://doi.org/10.1016/j.jalz.2019.06.1145 (2019).
41. Irva, H., Kauppinen, T., Pukkala, E. & Karjalainen, S. Cancer risk among workers exposed to arsenic and other inorganic compounds in the glass manufacturing industry. *Scand. J. Work Environ. & Health.* **18**(2), 74–82 (1992).
42. Poinen-Rughooputh, S., Rughooputh, M. S., Guo, Y., Rong, Y. & Chen, W. Occupational exposure to silica dust and risk of lung cancer: an updated meta-analysis of epidemiological studies. *BMC Public Health* **16**(1), 1137–1217. https://doi.org/10.1186/s12889-016-3791-5 (2016).
43. Emilie, R. et al. Asbestos exposure and the risk of lung cancer: A systematic review and meta-analysis. *Environ. Health Perspect.* **126**(4), 047002. https://doi.org/10.1289/EHP2735 (2018).

44. Lévêque, E. et al. Time-dependent effect of intensity of smoking and of occupational exposure to asbestos on the risk of lung cancer : results from the ICARE case–control study. *Occup. & Environ. Med.* **75**(8), 586–592 (2018).
45. Lindgren, A., Holmquist, C. & Jansson, B. Hypertension and the risk of lung cancer in men: A prospective cohort study. *J. Hypertens.* **21**(9), 1633–1638. https://doi.org/10.1097/00004872-200309000-00013 (2003).
46. Štěpánek, L., Ulčák, Z. & Konečný, D. The impact of secondhand smoke exposure on lung cancer mortality in non-smokers: Evidence from the Czech Republic. *Central Eur. J. Public Health.* **30**(3), 150–157. https://doi.org/10.21101/cejph.a6322 (2022).
47. Zhang, Y. et al. Alcohol use and lung cancer risk: A large-scale cohort study in Chinese populations. *Cancer Epidemiol. Biomark. & Prev.* **32**(7), 945–952. https://doi.org/10.1158/1055-9965.EPI-22-1129 (2023).
48. Cho, I., Han, K., Shin, D., Cho, M., Yoo, J., & Cho, J. Associations of variability in metabolic parameters with lung cancer: a nationwide population-based study Cancers, 13 8 https://doi-org.libproxy.txstate.edu/https://www.mdpi.com/2072-6694/13/8/1982. (2021)
49. Mbulo, L. et al. Secondhand smoke exposure at home among one billion children in 21 countries: Findings from the Global Adult Tobacco Survey (GATS). *Tob. Control* **25**, e95–e100 (2016).
50. Nadhiroh, S. R., Djokosujono, K. & Utari, D. M. The association between secondhand smoke exposure and growth outcomes of children: A systematic literature review. *Tob. Induc. Dis.* **18**, 12 (2020).
51. Ilaria, P. et al. Nutritional status and its role in lung cancer mortality: A comprehensive review and meta-analysis. *Nutr. & Cancer.* **72**(3), 408–416. https://doi.org/10.1080/01635581.2020.1723729 (2020).
52. García Lavandeira, M. et al. Alcohol consumption and the risk of lung cancer in never smokers: The SUN cohort study. *Eur. J. Public Health.* **28**(4), 636–642. https://doi.org/10.1093/eurpub/cky043 (2018).
53. García Lavandeira, J. A. et al. Alcohol consumption and lung cancer risk in never smokers: a pooled analysis of case-control studies. *Eur. J. Public Health* **28**(3), 521–527. https://doi.org/10.1093/eurpub/ckx196 (2018).
54. Bandera, E. V. et al. Alcohol consumption and lung cancer risk in the New York State cohort: A focus on never smokers and moderate drinkers. *Cancer Causes & Control.* **12**(10), 931–939. https://doi.org/10.1023/A:1012205215302 (2001).
55. Smith, A. H. et al. Lung, Bladder, and Kidney Cancer Mortality 40 Years After Arsenic Exposure Reduction. *JNCI : J. Natl. Cancer Inst.* **110**(3), 241–249. https://doi.org/10.1093/jnci/djx201 (2018).
56. Leiter, A., Veluswamy, R. R. & Wisnivesky, J. P. The global burden of lung cancer: current status and future trends. *Nat. Rev. Clin. Oncol.* **20**(9), 624–639. https://doi.org/10.1038/s41571-023-00798-3 (2023).
57. Wei, W. et al. Arsenic exposure and its joint effects with cigarette smoking and physical exercise on lung function impairment: Evidence from an occupational cohort study. *Environ. Res.* **196**, 110419. https://doi.org/10.1016/j.envres.2020.110419 (2021).
58. Oser, M. G., Niederst, M. J., Sequist, L. V. & Engelman, J. A. Transformation from non-small-cell lung cancer to small-cell lung cancer: molecular drivers and cells of origin. *Lancet Oncol.* **16**(4), e165–e172. https://doi.org/10.1016/S1470-2045(14)71180-5 (2015).
59. Abdik, H. Antineoplastic effects of erufosine on small cell and non-small cell lung cancer cells through induction of apoptosis and cell cycle arrest. *Mol. Biol. Rep.* **49**(4), 2963–2971. https://doi.org/10.1007/s11033-022-07117-6 (2022).

Acknowledgements
We extend our thanks to the two anonymous reviewers for their valuable suggestions and comments.

Author contributions
All authors reviewed the manuscript. Conceptualization and methodology, X.W., and J.Z.; investigation, resources, and data curation, X.W.; writing—original draft preparation, X.W.; writing, reviewing, and editing, and J.Z; supervision, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding
This research received no external funding.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92373-2.

Correspondence and requests for materials should be addressed to X.W. or J.Z.