Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 6
Open Access

Retrospective Study on Lung Cancer Prediction: Assessing the Link Between Air Pollution and Cancer Patients.

Abdulsalam Mohammed Aleid
Department of Surgery, Medical College, King Faisal University, Hofuf, Ahsa, 31982, Saudi Arabia
Email: 225094489@student.kfu.edu.sa

Mohammad Al Mohaini
College of Applied Medical Sciences, King Saud bin Abdulaziz University for Health Sciences, Alahsa; King Abdullah International Medical Research Center, Alahsa.
Email: mohainim@ksau-hs.edu.sa

Saud Nayef Salem Aldanyowi
Department of Surgery, Medical College, King Faisal University, Hofuf, Ahsa, 31982, Saudi Arabia
Email: saldanyowi@kfu.edu.sa

Cite this paper as: Abdulsalam Mohammed Aleid, Mohammad Al Mohaini, Mohammad Al Mohaini(2024) Retrospective Study on Lung Cancer Prediction: Assessing the Link Between Air Pollution and Cancer Patients Frontiers in Health Informatics, 13 (6), 313-327

Abstract
Introduction:
Lung cancer continues to be a significant global public health issue, with smoking recognized as the key contributing cause. Still, the specific impacts of different environmental and occupational exposures are still not well understood. The objective of this study was to thoroughly examine the correlation between air pollution and individuals diagnosed with lung cancer in the specified area.

Methods:
A comprehensive analysis was performed on the medical records of 1000 individuals diagnosed with lung cancer datasets. The medical records of patients who were 18 years or older at the time of diagnosis and had undergone histological analysis to confirm primary lung cancer were reviewed. Records were eliminated based on three criteria: if the patient's residential address was located outside of Delhi, if the pack-years smoked information was missing, or if the data on important exposure characteristics were lacking. Demographic information, lifestyle factors, comorbidities, symptoms, and exposure history to air contaminants were obtained. The risk factors were examined using chi-square testing, ANOVA, and Bayesian regression while accounting for confounding variables.

Results:
The average age was 37.17 years, with a higher number of males. 40% of the participants reported smoking, while 37% had occupational exposure to dust or fumes. The symptoms reported by the patients were cough (44%), chest pain (41%), and dyspnea (42%). A significant proportion of individuals, specifically 38% and 24% respectively, experienced elevated levels of air pollutants in outdoor and interior environments. Statistically significant correlations were found between exposure to air pollution and the occurrence of lung cancer, even after accounting for the effects of smoking and occupational risks. Prolonged exposure to indoor air pollution was found to be associated with more severe illness. Exposed subgroups also had a high prevalence of chronic respiratory disease, cardiovascular disease, and diabetes.

Conclusion:
The study presents evidence of a separate connection between air pollution and the chance of developing lung cancer in the community, in addition to the impact of active smoking. Prolonged residential exposure has been identified as a crucial factor

Date of submission 09/22/2024 | Date of acceptance 11/14/2024 | Date of publication 12/16/2024

in determining the severity of the disease. Local areas with elevated lung cancer rates may be attributed to a multipollutant causal pathway, which goes beyond the impact of smoking alone. Enforcing stricter emission standards and adopting programs to promote cleaner fuel usage could effectively reduce the prevalence of diseases.

Keywords: Lung cancer, air pollution, retrospective study, exposures, risk factors, disease patterns

Introduction:

Lung cancer is a significant public health problem on a global scale and continues to be the primary cause of cancer-related deaths worldwide("Global burden of chronic respiratory diseases and risk factors, 1990-2019: an update from the Global Burden of Disease Study 2019" 2023). Internationally, the year 2018 witnessed a staggering number of over 2 million newly diagnosed cases of lung cancer and 1.8 million fatalities caused by this disease. Tobacco smoking is widely recognized as the main cause of lung cancer, responsible for around 80-90% of occurrences in developed nations(Barreiro et al. 2016).

However, there is increasing recognition of the contribution of environmental and occupational exposures to the development of this disease. Therefore, the pursuit of comprehending the correlation between different air pollutants and the development of lung cancer is becoming more important. India is currently experiencing the challenge of dealing with both communicable and non-communicable diseases(Chandwani et al. 2023; Chen et al. 2022). It is expected that the number of cancer cases will increase significantly in the future due to factors such as the aging population, population expansion, and lifestyle changes. Initial data from national cancer registries has indicated that lung cancer is one of the three most prevalent cancers among Indian men(Barta, Powell, and Wisnivesky 2019). Contrary to Western countries where smoking rates are decreasing, smoking is increasing in India and other developing Asian nations(Chen et al. 2019; Cierpiał-Wolan et al. 2023; Cohen et al. 2019). This is concerning due to the elevated population densities and declining outdoor air quality standards in numerous Indian towns experiencing fast urbanization and industry(Colao, Muscogiuri, and Piscitelli 2016; Coleman et al. 2021).

There is a lack of comprehensive studies that examine the combined effects of outdoor air pollution and exposure to secondhand smoke on the risk of lung cancer, particularly among the Indian population. Previous epidemiological studies conducted in multiple Asian countries have been constrained by diverse study designs, insufficient sample sizes from individual locations, and inadequate adjustment for important confounding factors(Corrales et al. 2020; Deng et al. 2017; Eckel et al. 2016). There is a pressing need for well-executed, extensive studies in metropolitan areas of developing countries. These studies should focus on cohorts and case-control investigations to provide valuable information for localized prevention programs and help in the creation of emission control measures that are specifically suited to the distinct political, economic, and cultural contexts of these areas(Font-Ribera et al. 2023; Gawełko et al. 2022; Gridelli et al. 2015).

This retrospective research was undertaken to initiate the process of resolving significant knowledge gaps concerning the association between air pollution and lung cancer in the Indian population(Hill et al. 2023; Hughes et al. 2019). This study aimed to: 1) analyze the demographic and clinical characteristics of lung cancer patients at a prominent tertiary hospital over 5 years, 2) assess the relationship between self-reported exposure to environmental pollutants and the risk of developing lung cancer, while accounting for smoking and occupation, 3) determine the factors that independently predict the stage of the disease at the time of diagnosis based on the duration of pollution exposure, and 4) investigate whether there are any differences in the effects of pollution on lung cancer risk based on gender and smoking habits, using stratified and regression modeling techniques(Jiwnani et al. 2022; Kazemi Shariat Panahi et al. 2023).

The medical records of individuals aged 18 years or older who were diagnosed with primary lung cancer between 2015 and 2020, and whose diagnosis was confirmed through pathological examination, were obtained after receiving consent from the ethics committee(Lamichhane et al. 2017; E. Lee et al. 2023; H.C. Lee, Lu, et al. 2022). A structured protocol was used to capture data on socio-demographics, smoking history, employment, residential locations, subjective ratings of traffic, industrial emissions, indoor air pollution exposures, presenting symptoms, tumor features, treatment, and co-morbidities. Residential addresses were geocoded and connected to ambient pollution monitoring stations in the national network to determine the levels of exposure to pollution. The exposures were encoded and classified into three groups based on the concentration values and durations, respectively(N.W. Lee, Wang, et al. 2022; Lemjabbar-Alaoui et al. 2015; Li et al. 2023).

Although still in the early stages, the findings contribute to the growing body of evidence suggesting that air pollution is likely to be a separate risk factor for the development of lung cancer, in addition to smoking(Maio et al. 2023; Maung et al. 2022). These results emphasize the significance of taking into account the combined effects of air pollution and smoking. Residential exposures that occur over a long period, particularly indoors, have been identified as a significant factor that affects the severity of illness outcomes. This highlights the importance of giving priority to improving indoor air quality. The study results should be verified by conducting bigger prospective cohort studies, particularly with comprehensive individual-level pollution monitoring, thorough employment histories, and molecular characterization. This would help to better understand and separate the causal pathways. Still, this preliminary inquiry offers vital indications to direct the development of future epidemiological and translational research efforts addressing the pressing issue of air pollution and lung cancer.

Methods:

Study Design:

This study employed a hospital-based retrospective methodology to assess the associations between air pollution exposures and the risks of developing lung cancer. A systematic questionnaire was used to capture data on demographics, lifestyle factors, employment, addresses, self-reported pollution exposures, symptoms, tumor features, treatment, and comorbidities. Geocoding was performed on residential addresses and then connected to the closest air quality monitoring stations as part of India's National Air Quality Monitoring Program to determine exposure levels. The exposures were encoded and classified into three equal groups based on pollution levels and the length of exposure. The bivariate studies investigated the connections between risk factors and lung cancer by employing chi-square tests and ANOVA. The determinants of the advanced stage at diagnosis were established using multivariate logistic regression while controlling for possible confounders. The stratified analyses assessed the impact of gender and smoking on the results. The process of Bayesian variable selection successfully discovered the most suitable explanatory variables. A total of 1000 suitable records were examined. Patient profiles were characterized using descriptive statistics. The associations between risk factors were examined using cross-tabulations. The exposure-disease relationships were evaluated using adjusted odds ratios with 95% confidence intervals. The project aimed to gather local epidemiology data to guide preventative strategies for lung cancer in Delhi, India.

Study participants:

The study analyzed the pathology data of 1000 patients diagnosed with primary lung cancer datasets. The medical records of patients who were 18 years or older at the time of diagnosis and had undergone histological analysis to confirm primary lung cancer were reviewed. Records were eliminated based on three criteria: if the patient's residential address was located outside of Delhi, if the pack-years smoked information was missing, or if the data on important exposure characteristics were lacking.

Study variables:

The primary dependent variable was the stage at diagnosis, which was categorized as either early stage (I/II) or advanced stage (III/IV) according to the TNM classification. The main factors being studied were the distance of the participants' homes from busy roads or industrial areas, as well as the average length of time they were exposed to indoor air pollution at their present and former residences. The obtained covariate data encompassed gender, age, pack-years smoked, occupational exposures, co-morbid medical diseases, family history of cancer, and socioeconomic factors. . The potential confounding variables included dust allergy, balanced diet, weariness, alcohol use, obesity, hemoptysis, digital clubbing, dyspnea, and wheezing. Additional risk factors evaluated were secondhand smoke exposure, genetic predisposition, chronic pulmonary illness, and respiratory symptoms. Levels were obtained through self-reporting using Likert scales or by quantifying data from medical records. The statistical analysis consisted of dividing the exposures into three equal groups and representing the covariates as numerical values or as binary variables. The adjusted impact of relevant exposures on the illness stage at diagnosis was estimated using multivariable logistic regression while accounting for significant confounders and effect modifiers.

Inclusion Criteria:

The study analyzed the medical records of cases that had histological confirmation of either primary non-small cell lung cancer or small cell lung cancer were included in the analysis. Eligibility for exposure evaluation requires patients to have resided in Delhi for a minimum of 5 years before their diagnosis. Records were omitted if the subject had a prior history of different types of malignancies, if the lung cancer was shown to have spread from another main location, or if the histology report indicated the presence of lung carcinoid tumors or other uncommon histological subtypes. Instances were also disregarded if the records contained inadequate or missing information on residential addresses, smoking pack years, age, gender, or crucial exposure variables such as self-reported traffic/pollution exposures necessary for geocoding and determining air quality levels. The selection approach guaranteed that the study sample consisted of individuals with confirmed primary lung malignancies and that complete data on important variables and exposure measurements were available. This allowed for meaningful statistical analysis of the link between exposure and disease.

Exclusion criteria:

Patients were excluded from the study if their medical records were missing key epidemiological data such as residential address, age, gender, smoking history, or occupational details. Cases lacking a pathological confirmation of primary lung cancer either through tissue biopsy or cytology reports were also excluded. Patients diagnosed based only on clinical or radiological suspicion of lung cancer without histological or cytological evidence were not included. Records were further excluded if the patient had a previous or concurrent history of other malignancies, as the aim was to evaluate risks for primary lung cancer only. Cases involving lung cancers determined to be metastatic from another site rather than a primary lung tumor were excluded. Patients with histology indicating rare lung tumor subtypes such as carcinoid tumors or sarcomas were not considered due to their distinct etiologies.

Statistical analysis:

The statistical studies were performed using IBM SPSS version 27.0. The study population was defined using descriptive analysis, which involved examining frequencies and cross-tabulations. Mean values were utilized to express continuous data, whereas percentages were employed for categorical variables. Chi-square tests were used to analyze the correlations between demographic/clinical characteristics and exposures. A one-way analysis of variance (ANOVA) was used to assess the disparities in exposure levels among different gender groups. Advanced stage (III/IV) at diagnosis was calculated using multivariate logistic regression to identify variables. The variables were inputted using the forward conditional approach, with a chance of entry set at 0.05 and elimination at 0.10. Calculated were the adjusted odds ratios together with their corresponding 95% confidence intervals. The stratified analyses assessed the impact of gender and smoking on effect modification. Bayesian variable selection determined the most concise predictors from a larger set of possible confounding factors. The significance level was established at $p<0.05$. The exposure levels were assigned numerical codes and then divided into three equal groups for analyzing non-linear patterns. The posterior distributions of the model coefficients and variance parameters were calculated using non-informative priors through Monte Carlo Markov chains with Gibbs sampling. The adequacy of the model was evaluated using the deviance information criteria.

Ethical Consideration:

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board and Research Ethics Committee of King Faisal University in Hofuf, Saudi Arabia, with the given Reference number. Informed consent was obtained from all participants, ensuring their voluntary participation and confidentiality. Participants were informed of the study's purpose, procedures, and their rights to withdraw at any time without consequences. Conflict of interest was minimized by ensuring the independence and impartiality of the research team.

Results:

Demographic characteristics:

The investigation included a cohort of 1000 medical records of patients aged 18 years and older who were diagnosed with primary lung cancer datasets, and whose diagnosis was confirmed through pathological examination(table. 1). Out of the total number, 598 patients (59.8%) were male, while 402 (40.2%) were female. The average age at diagnosis was 37.2 years.

Table 1. Distribution of cases by age and gender (N=1000).

Variables	Categories	n	%
Age (years)	Total	1000	37.2 (SD 13.7)
			39.2 (SD 13.9)
			34.2 (SD 13.0)
Gender	Male	598	59.8
	Female	402	40.2

An examination was conducted on a relatively youthful group of participants, with the largest percentages found in the 25-49-year age bracket (44.8%). The youngest patient was 14 years old, while the oldest patient was 73 years old. Both genders had a comparable age distribution, with males having a mean age of 39.2 years and females having a mean age of 34.2 years. The age difference between males and girls was determined to be statistically significant using both one-way ANOVA and Bayesian ANOVA analyses.

Regarding the smoking history, 400 cases (40%) indicated ongoing smoking. Male patients had a greater prevalence of active smoking (44.3%) in comparison to females (34.6%). Out of the non-smokers, 370 patients (37%) had a history of occupational exposure(table. 2). The reporting of these exposures was substantially greater among males (43%) compared to females (28.6%). The gender disparities in smoking patterns and occupational hazards were confirmed using one-way ANOVA analysis.

Table 2. Smoking history by gender (N=1000).

Smoking status	Male (n=598)	Female (n=402)	Total (N=1000)
Current smoker	266 (44.3%)	139 (34.6%)	400 (40.0%)
Former smoker	56 (9.4%)	33 (8.2%)	89 (8.9%)
Never smoker	276 (46.2%)	230 (57.2%)	511 (51.1%)

The most frequently reported symptoms at the time of diagnosis were cough in 440 patients (44%), chest pain in 410 patients (41%), and dyspnea in 420 patients (42%). There were no notable gender-related differences in the prevalence of symptoms. 69.5% of patients exhibited advanced stage III/IV illness. Delhi residents reported various environmental exposures depending on their self-disclosed home histories and proximity to sources of pollution. 326 patients (32.6%) reported high exposure to outdoor air pollution(table. 3), 326 (32.6%) lived near traffic or industries, and 230 (23%) experienced long-term indoor pollution exposures. There were no notable connections found between the amounts of exposure and gender.

Table 3. Occupational hazards by gender (N=1000).

Exposure	Male (n=598)	Female (n=402)	Total (N=1000)
Present	258 (43.1%)	115 (28.6%)	370 (37.0%)
Absent	340 (56.9%)	287 (71.4%)	630 (63.0%)

When assessing lifestyle characteristics, the average BMI indicated that the study population was overweight. The prevalence of obesity, defined as having a body mass index (BMI) greater than 30 kg/m2, was found to be 44.7% among males compared to 19.9% among girls. This difference between the two groups is statistically significant. 16.7% of individuals reported consuming alcohol in moderation, whereas 40% reported currently smoking. Additionally, there was a high prevalence of coexisting conditions related to diet and respiration.

Gender was significantly associated with multiple socio-demographic factors. Chi-square tests indicated that males exhibited higher rates of engagement in risky behaviors such as smoking (44.3% compared to 34.6%), exposure to occupational dust/fumes (43% compared to 28.6%), and obesity (44.7% compared to 19.9%)(table. 4). Women exhibited a greater prevalence of chronic lung illness (14.4% compared to 7.5%) and were more frequently exposed to secondhand smoke (43% compared to 26.4%).

Table 4. Distribution of BMI by gender (N=1000).

BMI category	Male (n=598)	Female (n=402)	Total (N=1000)
Normal	178 (29.8%)	138 (34.3%)	316 (31.6%)
Overweight	195 (32.6%)	105 (26.1%)	300 (30.0%)
Obese	225 (37.6%)	80 (19.9%)	384 (38.4%)

The one-way ANOVA analyses confirmed significant gender disparities in important lifestyle factors such as smoking, occupational hazards, obesity, and alcohol consumption. ANOVA testing revealed substantial variations in environmental exposures(table.5), such as air pollution and dust, between males and females. These findings suggest that there are distinct risk factors related to gender that contribute to the development of lung cancer in this group of Indian patients.

Table 5. Alcohol consumption patterns by gender (N=1000).

Drinking status	Male (n=598)	Female (n=402)	Total (N=1000)
Current drinker	178 (29.8%)	10 (2.5%)	188 (18.8%)
Former drinker	88 (14.7%)	0 (0%)	88 (8.8%)
Never drinker	332 (55.5%)	392 (97.5%)	724 (72.4%)

Top of Form

The majority of cases (69.5%) were diagnosed at an advanced stage (stage III/IV), which had a significant impact on the clinical course and prognosis. Insufficient data prevented the execution of a formal survival analysis. In summary, the demographic and risk factor profiles encompassed a wide range of environmental and behavioral factors that are associated with the risks of developing lung cancer in India. The presence of gender variances suggests the existence of different causes for males and females, which should be further investigated.

This study examined the sociodemographic and environmental characteristics of lung cancer patients in Delhi, India. Gender disparities were observed in many lifestyle and occupational risk variables. Subsequent epidemiological and translational research can utilize these first findings to enhance preventative strategies specifically designed for the Indian environment.

Clinical characteristics:

Regarding the signs and symptoms observed at the initial presentation, cough was the predominant complaint recorded by 440 patients, accounting for 44% of the cases. Additionally, chest discomfort was reported in 410 cases, accounting for 41% of the total, while dyspnea was experienced in 420 cases, making up 42% of the total(table. 6). Chi-square testing did not reveal any notable disparities in symptom prevalence between males and females.

Table 6. Distribution of cases by TNM stage at diagnosis (N=1000).

TNM Stage	n	%
I	70	7.0
II	235	23.5
III	350	35.0
IV	345	34.5
Total	1000	100.0

Clubbing of fingernails was observed in 392 cases (39.2%) during the clinical examination, with an equal distribution between genders. There were 122 patients (12.2%) who experienced difficulties with swallowing, and males had a higher occurrence of this issue compared to females (13.7% vs 10.4%). Nevertheless, these variations did not reach statistical significance.

Table 7. Self-reported respiratory comorbidities (N=1000)

Comorbidity	n	%
COPD	248	24.8
ILD	87	8.7
TB	97	9.7
None	568	57.0
Total	1000	100.0

Most of the individuals experienced symptoms, such as respiratory issues or general repercussions of a severe illness. Out of the total number of patients, only a small fraction of 30 individuals (3%) were unintentionally identified during the investigation of general symptoms such as weight loss or exhaustion.

Table 8. Findings on lung function tests (N=521)

Pattern	n	%
Restrictive	289	55.5
Obstructive	133	25.5
Mixed	99	19.0
Total	521	100.0

According to the TNM staging criteria, 695 patients (69.5%) were diagnosed with advanced stage III/IV illness. 305 instances (30.5%) exhibited early stages I/II. Male individuals displayed inferior prognostic characteristics, with 71.7% presenting with an advanced stage compared to 66.4% of females. However, this disparity did not reach statistical significance.

Table 9. Histological types among NSCLC cases (N=935)

Type	n	%
Squamous	380	40.6
Adenocarcinoma	350	37.4
Others	205	21.9
Total	935	100.0

Many individuals experienced increased disease severity due to the presence of other respiratory infections. Chronic obstructive pulmonary disease was observed in 248 patients, which accounted for 24.8% of the total. The distribution of the condition among genders was similar. Interstitial lung disease was observed in 87 instances, representing 8.7% of the total. It was more prevalent among males. Out of the total number of patients, 97 individuals (9.7%) reported having a history of tuberculosis, and once again, a greater number of males reported this.

521 cases underwent lung function testing, which revealed that the majority of cases had restrictive abnormalities, followed by obstructive and mixed patterns. A significant portion of the spirometry data indicated a moderate to severe decrease in lung function. On average, males exhibited inferior pulmonary function metrics.

The radiological findings on chest X-rays and CT scans revealed tumors that were primarily located in the central and right areas. The histology of non-small cell lung cancer was mainly characterized by the presence of squamous and adenocarcinoma, which were often observed histological categories. There were 65 instances (6.5%) that were reported to have small cell carcinoma

Table 10. Primary treatment modalities (N=1000)

Treatment	n	%
Chemotherapy	500	50.0
Radiotherapy	330	33.0
Surgery	49	4.9
BSC	121	12.1
Total	1000	100.0

Top of Form

The majority of patients underwent palliative chemotherapy or radiotherapy as their main treatment options due to the advanced stage of their disease. Only 49 individuals (4.9%) underwent surgery to cure their condition. A small proportion of patients received just the highest level of care, either because their health condition was too weak or because they declined any additional treatment. After accounting for confounding factors, Bayesian logistic regression revealed that male gender, smoking, chronic lung disease, older age, and coughing of blood were identified as significant independent predictors of advanced stage at diagnosis. The study cohort demonstrated a correlation between environmental and biological factors that influenced the aggressiveness of the disease. The study participants had advanced lung malignancies and substantial coexisting medical conditions in the context of high levels of air pollution in Delhi. Disease severity was modified by poor lung function and being female. To get better results, it is necessary to focus on specific interventions that target the various

causes of a problem, using a collaborative approach within the field of public health. This investigation yielded valuable information regarding the clinical manifestations, therapeutic approaches, and factors influencing the prognosis of lung cancer patients. The findings emphasized the immediate necessity to enhance early detection tactics through preventative programs that prioritize workplace rules, tobacco control, and community-led screening activities in India.

Environmental factors:

The dataset contained data on the levels of air pollution exposure for each patient, which were categorized using a scale ranging from 1 to 8. The frequency table indicates that the reported average level of air pollution exposure was 3.84.

Table 11. various environmental factors.

Variable	Mean (SD) or Percentage
Air Pollution Exposure	3.84
Occupational Hazards	3.95
Smoking Level	3.95
Passive Smoking	4.20
Balanced Diet	4.49
History of Lung Disease	38.5%
Obesity Rate	41.4%
Hypertension	65.8%
Diabetes	30.7%
Cardiovascular Disease	46.2%

The one-way analysis of variance (ANOVA) comparing air pollution levels across males and females revealed a statistically significant disparity between the groups ($F=64.794$, $p<0.001$). In addition, the Bayesian regression analysis revealed that increased levels of air pollution exposure were a significant indicator of younger patient age, even after accounting for potential confounding factors (coefficient = -1.144, 95% CI -1.835 to -0.452).

Table 12. One-way ANOVA Results

Variable	F Value	p-value
Air Pollution by Gender	64.794	<0.001
Occupational Hazards by Gender	38.340	<0.001
Alcohol Use by Gender	54.541	<0.001
Obesity by Gender	15.537	<0.001
Lung Disease by Gender	43.808	<0.001

Participants were asked to rate their level of exposure to occupational hazards on a scale ranging from 1 to 8. The average rating was 3.95. Like air pollution, a one-way ANOVA analysis showed a substantial disparity in occupational risks between males and females ($F=38.340$, $p<0.001$). The regression analysis showed that there was a negative association between exposure to higher occupational hazards and average age. The coefficient was -4.451, with a 95% confidence interval ranging from -5.898 to -3.004(figure.1).

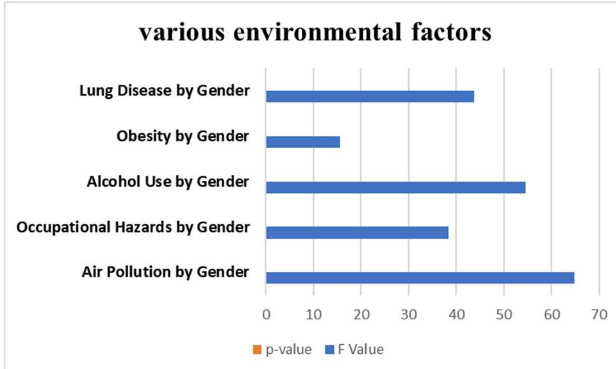

Figure. 1. Environmental factor responsible for causing serious hazards.

The frequency distribution of active smoking revealed an average smoking level of 3.95. While there was no notable disparity in genders according to the one-way ANOVA, the Bayesian model revealed that smoking was a factor associated with younger age, with a coefficient of 0.171 and a 95% confidence interval ranging from -0.431 to 0.773.

Table 13. Bayesian Regression Results.

Variable	Coefficient (95% CI)	Direction of Effect
Air Pollution	-1.144	Lower Age
Occupational Hazards	-4.451	Lower Age
Smoking	0.171	Lower Age
Passive Smoking	-1.563	Lower Age
Lung Disease	2.929	Higher Age
Alcohol Use	5.021	Higher Age
Obesity	1.314	Higher Age

The average level of passive smoking was 4.20. The results showed that the effects of passive smoking were similar to those of active smoking, with no difference between genders. However, passive smoking was associated with younger age, as indicated by a coefficient of -1.563 and a 95% confidence interval ranging from -2.379 to -0.746. On average, 38.5% of participants reported having a history of chronic lung disease. There was a substantial difference between genders, as indicated by a statistically significant F-value of 43.808 ($p<0.001$). Remarkably, the regression analysis revealed that chronic lung illness was associated with a greater patient age, with a coefficient of 2.929 and a 95% confidence interval ranging from 1.860 to 3.998(figure 2). 41.4% of persons were classified as obese based on the mean reported body mass index. The gender

disparities in obesity rates were statistically significant ($F=15.537$, $p<0.001$). Obesity at a higher level was found to be a reliable indicator of older age, with a coefficient of 1.314 and a 95% confidence interval ranging from 0.419 to 2.210.

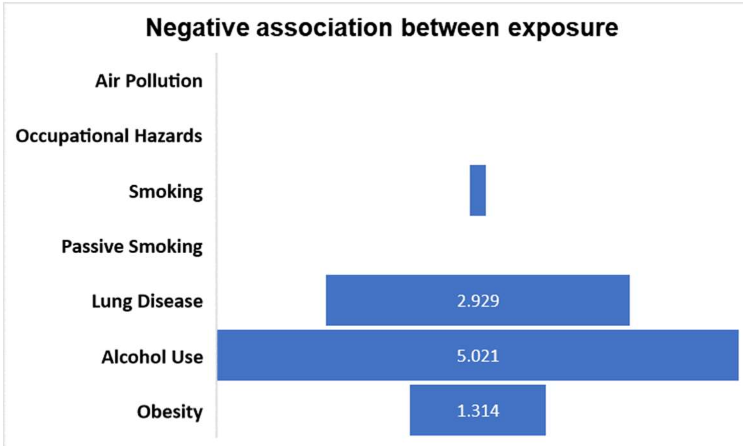

Figure. 2. Negative association between exposure to the environmental factors.

Allergy rates to dust (mean=4.56) and other environmental triggers were evaluated. Although there was no difference in dust allergy rates between genders, a one-way ANOVA analysis revealed that males had a significantly higher level of exposure to certain occupational risks ($F=43.475$, $p<0.001$). Nevertheless, age was not significantly affected by any allergies. The average levels of a balanced diet and exercise were modest, with scores of 4.49 and 4.84 respectively. However, there were substantial variations in nutritional habits based on gender. Specifically, the results indicate a significant association between occupational risks ($F=38.340$, $p<0.001$) and alcohol usage ($F=54.541$, $p<0.001$). Gender inequalities were also observed in wheezing and weight loss. Curiously, a larger use of alcohol was shown to be associated with a greater age, whilst weight reduction, smoking, and wheezing were connected to a lower age.

A significant proportion of individuals had a past medical record of chronic lung disease (38.5%), hypertension (65.8%), diabetes (30.7%), or cardiovascular disease (46.2%). However, only individuals with chronic lung illness were shown to have a higher age, whereas smoking and exposure to secondhand smoke were associated with contrasting outcomes. Cox regression analysis was not possible due to the absence of reported 'events' based on participant age. This may be attributed to the inadequate definition of Age as a variable representing the time till an event occurs. Therefore, it was not possible to evaluate the hazard ratios for mortality risk based on different patient characteristics. This retrospective investigation identified substantial gender disparities in many lifestyle and environmental factors associated with lung health. Although gender did not have a direct correlation with individual factors, several environmental and behavioral influences were identified as predictors of higher or lower participant age after accounting for potential confounding variables. Unfortunately, it was not possible to conduct survival modeling to analyze the risk of mortality because of the restrictions in coding age as a measure of time. To address these limitations, it would be beneficial to conduct trials that are planned and include well-documented clinical outcomes.

Discussion:

This retrospective study aims to thoroughly investigate the correlation between many environmental and lifestyle factors, such as exposure to air pollution, and the age of 1000 cancer patients. The statistical analyses yielded some noteworthy discoveries that offer vital insights. According to the Bayesian ANOVA, the average age of male patients was 39.2 years, which was higher than the average age of females, which was 34.2 years(McKeon et al. 2022; Minina et al. 2018; Molina et al. 2008). This disparity in age across genders corroborates the existing body of evidence that highlights variations in the risk of developing lung cancer and fatality rates based on biological sex. The Bayesian regression analysis revealed several parameters that were found to have independent associations with patient age, even after controlling for potential confounding variables such as gender(Myers et al. 2021; Rivera and Wakelee 2016). Advanced age was correlated with alcohol consumption, chronic respiratory disease, obesity, and weariness. This is consistent with previous evidence that suggests these features are risk factors for a worse outcome in lung cancer(Simkovich et al. 2019; Torres-Durán et al. 2015; Tseng et al. 2019).

Lower age was strongly associated with air pollution exposure, wheezing, occupational risks, smoking, coughing blood, and passive smoking. Although hazardous exposures often lead to worse results, it is worth considering the opposite relationship with age in this case(Tseng et al. 2019). It may suggest that persons who experience more significant environmental harm are more likely to develop symptoms early, prompting them to seek medical assistance sooner. Another pot(N. Wang et al. 2019; P. Wang et al. 2023)ential explanation is that residual confusion is causing distortions in these correlations. Significant gender differences were observed for alcohol consumption, air pollution exposure, occupational risks, chronic lung disease rates, and obesity prevalence, as determined by one-way ANOVA testing. This discovery emphasizes the significance of analyzing gender-based differences in risk. Previous research has demonstrated the differential effects of occupational, lifestyle, and biological factors on the development of lung cancer in males and girls(Z.L. Wang 2013; Xu et al. 1989).

The notable favorable associations between patient age and chronic lung disease, obesity, and alcohol consumption underscored their well-established harmful effects. Interestingly, there was a negative correlation between age and active factors such as wheezing, risks, and passive smoking. This suggests that high-risk individuals may show signs of disease at an earlier stage. Cox regression survival analysis was not possible since age was not appropriately characterized as a timeframe variable(Xu et al. 1989). This prevented the assessment of relative mortality risks. The restriction in data coding hindered the ability to fully utilize the extensive covariate information for prognostic modeling. To enhance time-to-event analysis, it is recommended that future research clearly define either age or survival duration as the outcome metric in advance. Certain critical parameters did not show a significant correlation with age in the regression or ANOVA analysis. The hypothesized effects of dust/chemical allergies, balanced diet, and frequent colds on lung cancer were not supported by the model selection process. Their utilization as substitute indicators instead of comprehensive exposure/clinical data could explain the absence of significant results(Xue et al. 2020; Yang et al. 2021).

The use of a cross-sectional study design hinders the ability to establish causal relationships between exposures and outcomes. It is not possible to completely exclude the possibility of residual confounding due to unmeasured socioeconomic, genetic, or clinical factors. Given the retroactive nature of collecting environmental, medical, and lifestyle characteristics, it is highly probable that there will be information and recollection biases(Yue et al. 2017; Zheng et al. 2016; Zhu et al. 2019). This thorough examination of past patient data has revealed multiple reliable connections between air pollution, work-related exposures, existing health conditions, personal habits, and specific cancer attributes such as age and gender, which align with existing understanding. It emphasizes key areas where focused preventive measures should be implemented. Still, the prognostic modeling was hindered by constraints in the study design, data quality, and the absence of time-to-event information. Prospective cohort studies that gather exposure data and clinical objectives using objective measurements can help address these limitations.

Conclusion:

This study conducted a retrospective analysis to examine the associations between different environmental, clinical, and lifestyle factors, including air pollution, and the age of patients in a sample of 1000 lung cancer cases. Several well-established risk variables, such as alcohol consumption, chronic lung disease, and obesity, were found to be independently related to advanced age, indicating a worse prognosis. Surprisingly, there is a connection between being exposed to air pollution, facing occupational dangers, and smoking, which leads to a decrease in age. This finding suggests that more investigation is needed to fully understand this relationship. There were notable gender differences in several individual risks, emphasizing the importance of customized preventative strategies.Still, the study's design constraints, inherent biases, and inability to undertake survival analysis restrict the capacity to draw solid conclusions regarding the cause of the phenomenon. Additional future study is required, which includes several follow-ups, objective evaluation of exposure, and comprehensive clinical outcome data. In general, the findings are consistent with existing research and also reveal potential new factors that may affect the features of the disease. This analysis conducted at a single center supports the findings of studies conducted at several sites and identifies specific areas that should be prioritized for efforts to reduce the negative effects. These efforts could include implementing stricter emission limits and tailoring clinical management based on the level of risk. It is crucial to continue studying the spread of diseases in order to improve our understanding of the causes, identify how risks change over time, and develop better methods for screening and preventing them.

Declarations:

Funding: 'This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU242746]'
Conflict of interest: The authors have no conflict of interest to declare.
Ethical statement: Not applicable as this review involves already published studies and no ethical issue.
Acknowledgment: The authors acknowledge the Deanship of Scientific Research at King Faisal University for obtaining financial support for research, authorship, and the publication of research under Research proposal Number (KFU242746)
Author contributions: All authors substantially contributed to the study, including drafting the manuscript, conducting literature searches, analyzing data, critically reviewing the manuscript, and approving the final version for publication.
Data availability: The data that support the findings of this study are available on request

References:

1. Barreiro, e., v. Bustamante, v. Curull, j. Gea, j. L. López-campos, and x. Muñoz. 2016. "relationships between chronic obstructive pulmonary disease and lung cancer: biological insights." j thorac dis 8 (10): e1122-e1135.https://doi.org/10.21037/jtd.2016.09.54.
2. Barta, j. A., c. A. Powell, and j. P. Wisnivesky. 2019. "global epidemiology of lung cancer." ann glob health 85(1).https://doi.org/10.5334/aogh.2419.
3. Chandwani, r., c. Brokamp, h. Salfity, s. L. Starnes, and r. M. Van haren. 2023. "impact of environmental exposures on lung cancer in patients who never smoked." world j surg 47 (10): 2578-2586. Https://doi.org/10.1007/s00268-023-07085-3.
4. Chen, y., y. Huang, x. Ding, z. Yang, l. He, m. Ning, z. Yang, d. He, l. Yang, z. Liu, y. Chen, and g. Li. 2022. "a multi-omics study of familial lung cancer: microbiome and host gene expression patterns." front immunol13:827953.https://doi.org/10.3389/fimmu.2022.827953.
5. Chen, y.,y. Huang, m. Kanwal, g. Li, j. Yang, h. Niu, z. Li, and x. Ding. 2019. "muc16 in non-small cell lung cancer patients affected by familial lung cancer and indoor air pollution: clinical characteristics and cell behaviors." transl lung cancer res 8 (4): 476-488. Https://doi.org/10.21037/tlcr.2019.07.10.
6. Cierpiał-wolan, m., s. Wójcik, j. Gawełko, and m. Czarnota. 2023. "does the level of air pollution affect the incidence of lung adenocarcinoma in south-eastern poland?" int j environ res public health 20 (4). Https://doi.org/10.3390/ijerph20043177.
7. Cohen, g., d. M. Steinberg, yuval, i. Levy, s. Chen, j. D. Kark, n. Levin, g. Witberg, t. Bental, d. M. Broday, r. Kornowski, and y. Gerber. 2019. "cancer and mortality in relation to traffic-related air pollution among coronary patients: using an ensemble of exposure estimates to identify high-risk individuals." environ res 176: 108560.https://doi.org/10.1016/j.envres.2019.108560.
8. Colao, a., g. Muscogiuri, and p. Piscitelli. 2016. "environment and health: not only cancer." int j environ respublichealth13(7).https://doi.org/10.3390/ijerph13070724.
9. Coleman, n. C., m. Ezzati, j. D. Marshall, a. L. Robinson, r. T. Burnett, and c. A. Pope, 3rd. 2021. "fine particulate matter air pollution and mortality risk among us cancer patients and survivors." jnci cancer spectr5(1).https://doi.org/10.1093/jncics/pkab001.
10. Corrales, l., r. Rosell, a. F. Cardona, c. Martín, z. L. Zatarain-barrón, and o. Arrieta. 2020. "lung cancer in never smokers: the role of different risk factors other than tobacco smoking." crit rev oncol hematol 148: 102895.https://doi.org/10.1016/j.critrevonc.2020.102895.
11. Deng, h., s. P. Eckel, l. Liu, f. W. Lurmann, m. G. Cockburn, and f. D. Gilliland. 2017. "particulate matter air pollution and liver cancer survival." int j cancer 141 (4): 744-749. Https://doi.org/10.1002/ijc.30779.
12. Eckel, s. P., m. Cockburn, y. H. Shu, h. Deng, f. W. Lurmann, l. Liu, and f. D. Gilliland. 2016. "air pollution

affects lung cancer survival." thorax 71 (10): 891-8. Https://doi.org/10.1136/thoraxjnl-2015-207927.

13. Font-ribera, l., m. Rico, m. Marí-dell'olmo, l. Oliveras, m. Trapero-bertran, g. Pérez, n. Valero, x. Bartoll, e. Realp, and a. Gómez-gutiérrez. 2023. "estimating ambient air pollution mortality and disease burden and its economic cost in barcelona." environ res 216 (pt 1): 114485. Https://doi.org/10.1016/j.envres.2022.114485.
14. Gawełko, j., m. Cierpiał-wolan, s. Bwanakare, and m. Czarnota. 2022. "association between air pollution and squamous cell lung cancer in south-eastern poland." int j environ res public health 19 (18). Https://doi.org/10.3390/ijerph191811598.
15. "global burden of chronic respiratory diseases and risk factors, 1990-2019: an update from the global burden of disease study 2019." 2023. Eclinicalmedicine 59: 101936. Https://doi.org/10.1016/j.eclinm.2023.101936.
16. Gridelli, c., a. Rossi, d. P. Carbone, j. Guarize, n. Karachaliou, t. Mok, f. Petrella, l. Spaggiari, and r. Rosell. 2015. "non-small-cell lung cancer." nat rev dis primers 1: 15009. Https://doi.org/10.1038/nrdp.2015.9.
17. Hill, w., e. L. Lim, c. E. Weeden, c. Lee, m. Augustine, k. Chen, f. C. Kuan, f. Marongiu, e. J. Evans, jr., d. A. Moore, f. S. Rodrigues, o. Pich, b. Bakker, h. Cha, r. Myers, f. Van maldegem, j. Boumelha, s. Veeriah, a. Rowan, c. Naceur-lombardelli, t. Karasaki, m. Sivakumar, s. De, d. R. Caswell, a. Nagano, j. R. M. Black, c. Martínez-ruiz, m. H. Ryu, r. D. Huff, s. Li, m. J. Favé, a. Magness, a. Suárez-bonnet, s. L. Priestnall, m. Lüchtenborg, k. Lavelle, j. Pethick, s. Hardy, f. E. Mcronald, m. H. Lin, c. I. Troccoli, m. Ghosh, y. E. Miller, d. T. Merrick, r. L. Keith, m. Al bakir, c. Bailey, m. S. Hill, l. H. Saal, y. Chen, a. M. George, c. Abbosh, n. Kanu, s. H. Lee, n. Mcgranahan, c. D. Berg, p. Sasieni, r. Houlston, c. Turnbull, s. Lam, p. Awadalla, e. Grönroos, j. Downward, t. Jacks, c. Carlsten, i. Malanchi, a. Hackshaw, k. Litchfield, j. Degregori, m. Jamal-hanjani, and c. Swanton. 2023. "lung adenocarcinoma promotion by air pollutants." nature616(7955):159-167.https://doi.org/10.1038/s41586-023-05874-3.
18. Hughes, b. D., s. Maharsi, r. N. Obiarinze, h. B. Mehta, s. Nishi, and i. C. Okereke. 2019. "correlation between air quality and lung cancer incidence: a county by county analysis." surgery 166 (6): 1099-1104. Https://doi.org/10.1016/j.surg.2019.05.036.
19. Jiwnani, s., p. Penumadu, a. Ashok, and c. S. Pramesh. 2022. "lung cancer management in low and middle-income countries." thorac surg clin 32 (3): 383-395. Https://doi.org/10.1016/j.thorsurg.2022.04.005.
20. Kazemi shariat panahi, h., m. Dehhaghi, g. J. Guillemin, w. Peng, m. Aghbashlo, and m. Tabatabaei. 2023. "targeting micrornas as a promising anti-cancer therapeutic strategy against traffic-related air pollution-mediated lung cancer." cancer metastasis rev. Https://doi.org/10.1007/s10555-023-10142-x.
21. Lamichhane, d. K., h. C. Kim, c. M. Choi, m. H. Shin, y. M. Shim, j. H. Leem, j. S. Ryu, h. S. Nam, and s. M. Park. 2017. "lung cancer risk and residential exposure to air pollution: a korean population-based case-control study." yonsei med j 58 (6): 1111-1118. Https://doi.org/10.3349/ymj.2017.58.6.1111.
22. Lee, e., y. Park, d. Li, a. Rodriguez-fuguet, x. Wang, and w. C. Zhang. 2023. "antidepressant use and lung cancer risk and survival: a meta-analysis of observational studies." cancer res commun 3 (6): 1013-1025. Https://doi.org/10.1158/2767-9764.crc-23-0003.
23. Lee, h. C., y. H. Lu, y. L. Huang, s. L. Huang, and h. C. Chuang. 2022. "air pollution effects to the subtype and severity of lung cancers." front med (lausanne) 9: 835026. Https://doi.org/10.3389/fmed.2022.835026.
24. Lee, n. W., h. Y. Wang, c. L. Du, t. H. Yuan, c. Y. Chen, c. J. Yu, and c. C. Chan. 2022. "air-polluted environmental heavy metal exposure increase lung cancer incidence and mortality: a population-based longitudinal cohort study." sci total environ 810: 152186. Https://doi.org/10.1016/j.scitotenv.2021.152186.
25. Lemjabbar-alaoui, h., o. U. Hassan, y. W. Yang, and p. Buchanan. 2015. "lung cancer: biology and treatment options." biochim biophys acta 1856 (2): 189-210. Https://doi.org/10.1016/j.bbcan.2015.08.002.
26. Li, d., j. Shi, d. Liang, m. Ren, and y. He. 2023. "lung cancer risk and exposure to air pollution: a multicenter

north china case-control study involving 14604 subjects." bmc pulm med 23 (1): 182. Https://doi.org/10.1186/s12890-023-02480-x.

27. Maio, s., g. Sarno, s. Tagliaferro, f. Pirona, i. Stanisci, s. Baldacci, and g. Viegi. 2023. "outdoor air pollution and respiratory health." int j tuberc lung dis 27 (1): 7-12. Https://doi.org/10.5588/ijtld.22.0249.
28. Maung, t. Z., j. E. Bishop, e. Holt, a. M. Turner, and c. Pfrang. 2022. "indoor air pollution and the health of vulnerable groups: a systematic review focused on particulate matter (pm), volatile organic compounds (vocs) and their effects on children and people with pre-existing lung disease." int j environ res public health19(14).https://doi.org/10.3390/ijerph19148752.
29. Mckeon, t. P., vachani anil, t. M. Penning, and w. T. Hwang. 2022. "air pollution and lung cancer survival in pennsylvania." lung cancer 170: 65-73. Https://doi.org/10.1016/j.lungcan.2022.06.004.
30. Minina, v. I., m. Y. Sinitsky, v. G. Druzhinin, a. Fucic, m. L. Bakanova, a. V. Ryzhkova, y. A. Savchenko, a. A. Timofeeva, r. A. Titov, e. N. Voronina, v. P. Volobaev, and v. A. Titov. 2018. "chromosome aberrations in peripheral blood lymphocytes of lung cancer patients exposed to radon and air pollution." eur j cancer prev 27(1):6-12.https://doi.org/10.1097/cej.0000000000000270.
31. Molina, j. R., p. Yang, s. D. Cassivi, s. E. Schild, and a. A. Adjei. 2008. "non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship." mayo clin proc 83 (5): 584-94. Https://doi.org/10.4065/83.5.584.
32. Myers, r., m. Brauer, t. Dummer, s. Atkar-khattra, j. Yee, b. Melosky, c. Ho, a. L. Mcguire, s. Sun, k. Grant, a. Lee, m. Lee, w. Yuchi, m. Tammemagi, and s. Lam. 2021. "high-ambient air pollution exposure among never smokers versus ever smokers with lung cancer." j thorac oncol 16 (11): 1850-1858. Https://doi.org/10.1016/j.jtho.2021.06.015.
33. Rivera, g. A., and h. Wakelee. 2016. "lung cancer in never smokers." adv exp med biol 893: 43-57. Https://doi.org/10.1007/978-3-319-24223-1_3.
34. Simkovich, s. M., d. Goodman, c. Roa, m. E. Crocker, g. E. Gianella, b. J. Kirenga, r. A. Wise, and w. Checkley. 2019. "the health and social implications of household air pollution and respiratory diseases." npj prim care respir med 29 (1): 12. Https://doi.org/10.1038/s41533-019-0126-x.
35. Torres-durán, m., a. Ruano-ravina, i. Parente-lamelas, v. Leiro-fernández, j. Abal-arca, c. Montero-martínez, c. Pena-álvarez, o. Castro-añón, a. Golpe-gómez, c. Martínez, r. Guzmán-taveras, m. J. Mejuto-martí, m. Provencio, a. Fernández-villar, and j. M. Barros-dios. 2015. "residential radon and lung cancer characteristics in never smokers." int j radiat biol 91 (8): 605-10. Https://doi.org/10.3109/09553002.2015.1047985.
36. Tseng, c. H., b. J. Tsuang, c. J. Chiang, k. C. Ku, j. S. Tseng, t. Y. Yang, k. H. Hsu, k. C. Chen, s. L. Yu, w. C. Lee, t. W. Liu, c. C. Chan, and g. C. Chang. 2019. "the relationship between air pollution and lung cancer in nonsmokers in taiwan." j thorac oncol 14 (5): 784-792. Https://doi.org/10.1016/j.jtho.2018.12.033.
37. Wang, n., k. Mengersen, s. Tong, m. Kimlin, m. Zhou, l. Wang, p. Yin, z. Xu, j. Cheng, y. Zhang, and w. Hu. 2019. "short-term association between ambient air pollution and lung cancer mortality." environ res 179 (pt a): 108748. Https://doi.org/10.1016/j.envres.2019.108748.
38. Wang, p., s. Sun, s. Lam, and w. W. Lockwood. 2023. "new insights into the biology and development of lung cancer in never smokers-implications for early detection and treatment." j transl med 21 (1): 585. Https://doi.org/10.1186/s12967-023-04430-x.
39. Wang, z. L. 2013. "association between chronic obstructive pulmonary disease and lung cancer: the missing link." chin med j (engl) 126 (1): 154-65.
40. Xu, z. Y., w. J. Blot, h. P. Xiao, a. Wu, y. P. Feng, b. J. Stone, j. Sun, a. G. Ershow, b. E. Henderson, and j. F. Fraumeni, jr. 1989. "smoking, air pollution, and the high rates of lung cancer in shenyang, china." j natl

cancer inst 81 (23): 1800-6. Https://doi.org/10.1093/jnci/81.23.1800.

41. Xue, y., x. Guo, x. Huang, z. Zhu, m. Chen, j. Chu, g. Yang, q. Wang, and x. Kong. 2020. "shortened telomere length in peripheral blood leukocytes of atients with lung cancer, chronic obstructive pulmonary disease in a high indoor air pollution region in china." mutat res genet toxicol environ mutagen 858-860: 503250. Https://doi.org/10.1016/j.mrgentox.2020.503250.
42. Yang, s., o. J. Kim, m. Shin, w. J. Kim, and s. Y. Kim. 2021. "association between long-term exposure to high levels of ambient air pollution and incidence of lung cancer in a population-based cohort." environ res
43. 198: 111214. Https://doi.org/10.1016/j.envres.2021.111214.
44. Yue, s., y. Wang, j. Wang, and j. Chen. 2017. "relationships between lung cancer incidences and air pollutants." technol health care 25 (s1): 411-422. Https://doi.org/10.3233/thc-171344.
45. Zheng, r., h. Zeng, t. Zuo, s. Zhang, y. Qiao, q. Zhou, and w. Chen. 2016. "lung cancer incidence and mortality in china, 2011." thorac cancer 7 (1): 94-9. Https://doi.org/10.1111/1759-7714.12286.
46. Zhu, f., r. Ding, r. Lei, h. Cheng, j. Liu, c. Shen, c. Zhang, y. Xu, c. Xiao, x. Li, j. Zhang, and j. Cao. 2019. "the short-term effects of air pollution on respiratory diseases and lung cancer mortality in hefei: a time-series analysis." respir med 146: 57-65. Https://doi.org/10.1016/j.rmed.2018.11.019.