



OPEN

## Multiple air pollutant exposure and lung cancer in Tehran, Iran

Zahra Khorrami<sup>1</sup>, Mohsen Pourkhosravani<sup>2</sup>, Maysam Rezapour<sup>3</sup>, Koorosh Etemad<sup>4</sup>, Seyed Mahmood Taghavi-Shahri<sup>5</sup>, Nino Künzli<sup>6,7</sup>, Heresh Amini<sup>8,9</sup> & Narges Khanjani<sup>10,11,12</sup>✉

Lung cancer is the most rapidly increasing malignancy worldwide with an estimated 2.1 million cancer cases in the latest, 2018 World Health Organization (WHO) report. The objective of this study was to investigate the association of air pollution and lung cancer, in Tehran, Iran. Residential area information of the latest registered lung cancer cases that were diagnosed between 2014 and 2016 (N=1,850) were inquired from the population-based cancer registry of Tehran. Long-term average exposure to PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub>, benzene, toluene, ethylbenzene, *m*-xylene, *p*-xylene, *o*-xylene (BTEX), and BTEX in 22 districts of Tehran were estimated using land use regression models. Latent profile analysis (LPA) was used to generate multi-pollutant exposure profiles. Negative binomial regression analysis was used to examine the association between air pollutants and lung cancer incidence. The districts with higher concentrations for all pollutants were mostly in downtown and around the railway station. Districts with a higher concentration for NO<sub>x</sub> (IRR = 1.05, for each 10 unit increase in air pollutant), benzene (IRR = 3.86), toluene (IRR = 1.50), ethylbenzene (IRR = 5.16), *p*-xylene (IRR = 9.41), *o*-xylene (IRR = 7.93), *m*-xylene (IRR = 2.63) and TBTEX (IRR = 1.21) were significantly associated with higher lung cancer incidence. Districts with a higher multiple air-pollution profile were also associated with more lung cancer incidence (IRR = 1.01). Our study shows a positive association between air pollution and lung cancer incidence. This association was stronger for, respectively, *p*-xylene, *o*-xylene, ethylbenzene, benzene, *m*-xylene and toluene.

Lung cancer was the most common cancer worldwide with an estimated 2.09 million cancer cases in the latest, 2018 World Health Organization (WHO) report. It ranked as the first most common cancer for men (31.5 per 100,000) and second most common for women (14.6 per 100,000) after breast cancer. Lung cancer contributed to 18.3% of all cancer deaths worldwide<sup>1</sup>. The age standardized incidence rate (ASR) of lung cancer in 2015, in Iran was 8.52 per 100,000 ranging from 12.37 per 100,000 in West Azerbaijan to 3.84 per 100,000 in Sistan and Baluchistan province; and in Tehran (the country capital) was 8.74 per 100,000, which was more than the incidence in the whole country (Iran)<sup>2</sup>.

Air pollution is likely to be a risk factor for lung cancer. According to WHO, about 24% of the global burden of disease and 23% of global deaths can be attributed to environmental factors<sup>1,3</sup>. Air pollution consists of a complex mixture of particles and gases<sup>4</sup>, in differing amounts, depending on the types of emission sources and atmospheric conditions. Typically, air pollution consists of a range of pollutants such as soot, particulate matter (PM) ≤ 10 μm (PM<sub>10</sub>), and PM ≤ 2.5 μm (PM<sub>2.5</sub>), nitrogen dioxide (NO<sub>2</sub>), sulfur dioxide (SO<sub>2</sub>), carbon monoxide (CO), ozone (O<sub>3</sub>) and volatile organic compounds (VOCs)<sup>5</sup>. The latest International Agency for Research on Cancer (IARC) report has mentioned PM<sub>2.5</sub> as a risk factor for lung cancer<sup>6</sup> and the ELAPSE study in Europe showed that long-term ambient PM<sub>2.5</sub> exposure may cause lung cancer even at concentrations lower than current European threshold values and WHO Air Quality Guidelines<sup>7</sup>.

<sup>1</sup>Neurology Research Center, Kerman University of Medical Sciences, Kerman, Iran. <sup>2</sup>Department of Geography and Urban Planning, Shahid Bahonar University of Kerman, Kerman, Iran. <sup>3</sup>Amol Faculty of Paramedical Sciences, Mazandaran University of Medical Sciences, Sari, Iran. <sup>4</sup>Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran. <sup>5</sup>Department of Epidemiology and Biostatistics, School of Public Health, Isfahan University of Medical Sciences, Isfahan, Iran. <sup>6</sup>Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland. <sup>7</sup>University of Basel, Basel, Switzerland. <sup>8</sup>Department of Public Health, Section of Environmental Health, University of Copenhagen, Copenhagen, Denmark. <sup>9</sup>Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>10</sup>Environmental Health Engineering Research Center, Kerman University of Medical Sciences, Kerman, Iran. <sup>11</sup>Monash Centre for Occupational & Environmental Health, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. <sup>12</sup>Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran. ✉email: n\_khanjani@kmu.ac.ir

Humans are simultaneously exposed to a complex mixture of air pollutants; therefore, many researchers have investigated a multiple-pollutant approach for assessing air pollution exposure<sup>8</sup>, because in single pollutant models, it is not clear if an observed association reflects the effect of the specific pollutant under study, or the effect of coinciding pollutants<sup>9</sup>. However, there is no consensus on the method used for measuring multiple ambient air pollutants simultaneously. Previous review studies have evaluated multiple pollutants by using methods that can be classified into three groups; dimension reduction, variable selection, and grouping of observations<sup>4,8,10,11</sup>. However, Caban-Martinez, et al.<sup>12</sup> and Kolpacoff, et al.<sup>13</sup> have used latent profile analysis (LPA) to identify subgroups of cancerous patients with different multi-pollution profiles. LPA is a probabilistic or model-based technique and is a variant of the traditional cluster analysis, which better handles outliers and unequal cluster sizes<sup>14</sup>. This method enables identifying possible unobservable subgroups, or latent classes in a population using a number of related observable variables. Using this method, complex relations between groups of risk factors and disease outcomes, such as cancer, which may not be best explained by a single pollutant model, can be better understood. It also reduces the dimensionality of exposure data and decreases the burden of multiple testing, while enhancing the power of statistical analysis<sup>13</sup>.

Knowledge about geographical patterns of multiple pollution helps policy makers to target high risk regions for more intense interventions. Air pollution exacerbates the health disparity among socioeconomic groups, because usually the poor socio-economic areas are more polluted<sup>15,16</sup>. The relation between air pollution exposure and health outcomes can also be theoretically modified by socioeconomic status, through causing differences in access to medical care and healthy diet, and also by biological factors, such as age and psychological stress. However, the hypothesis that residents with a low socioeconomic position face more severe consequences from air pollution is debated<sup>16</sup>.

In the current study, we aimed to examine the association between single and multiple ambient air pollutants and lung cancer incidence in Tehran, Iran.

## Methodology

**Research location.** This study was conducted in Tehran, a megacity which is the most populous city in Iran, with a residing population of about 9 million and a daytime population of over 14 million people. According to the World Population Review report, Tehran's 2020 population is now estimated to be 9,134,708. Tehran is the most populous city in Iran and Western Asia, and has the third-largest metropolitan area in the Middle East. It is ranked 24th in the world by the population of its metropolitan area. Tehran's area extend is about 730 km<sup>2</sup> and consists of 22 municipal districts with different concentrations of ambient air pollutants<sup>17</sup>.

**Data sources.** *Lung cancer data.* Totally, 1850 patients residing in Tehran were diagnosed with lung cancer (Trachea, Bronchus and Lung) between 2014 and 2016. The latest address of these patients' residence in Tehran was inquired from the Cancer Department of the Ministry of Health of Iran. The officials of the cancer registry claimed that the recorded addresses are more than 90% accurate.

The geographical coordinates (longitude and latitude) of the participants' residential addresses were determined according to the address of the patients' residence and was marked on the GIS map of Tehran.

*Exposure assessment.* The annual mean concentrations of PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub>, benzene, toluene, ethylbenzene, *m*-xylene, *p*-xylene, *o*-xylene (BTEX), and total BTEX in the 22 districts of Tehran were inquired from previously developed land use regression (LUR) models. The LUR models for PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub> and NO<sub>x</sub> in Tehran were developed based on measurements conducted at 23 sites in Tehran in 2010<sup>18,19</sup>. The models for volatile organic compounds (VOCs) were developed based on measurements at 179 sampling sites from April 2015 to May 2016<sup>20,21</sup>.

*Confounding covariates.* Population-based data was extracted from the Urban Health Equity Assessment and Response Tool (Urban HEART-2), which has been conducted in 22 districts of Tehran and is a data repository that collected many district-level variables, such as population density, per capita urban green space, smoking rates and life expectancy in 2011. A detailed description of the Urban HEART-2 study can be found elsewhere<sup>22</sup>.

The socio-economic development situation of the 22 districts of Tehran was extracted from a study conducted by Sadeghi et al.<sup>23</sup>. In brief, sixteen economic and social indicators were incorporated to estimate the level of development in the 22 districts of Tehran, based on Exploratory Factor Analysis (EFA), Principal Component Analysis (PCA). This multivariate statistical technique is used to reduce the number of variables in a dataset into a smaller number of "dimensions" that explains most variations in the dataset using a few estimated substitutional latent variables<sup>23</sup>.

In Sadeghi et al.'s study, the variables used for estimating the social dimension included adult literacy level (among the 30–59 year olds), elderly literacy level (60 years and older), the proportion of university graduates in the total population, the proportion of the males with university education, the proportion of females with university education, and the percent of population that uses the internet. The variables used for estimating the economic part, included women's economic participation rate, the proportion of employees with high-rank jobs, the proportion of households with cars, the proportion of households with computers, indicators of household access to public facilities, the ratio of households who own their home, the proportion of homes with civil standards, the average price per square meter of residential building land, the average selling price per square meter of residential building area and the average monthly rent per square meter of residential area for each district. A detailed description of these variables can be found elsewhere<sup>23</sup>. A higher socio-economic development score showed a higher socio-economic level. This variable did not have a specific unit, its minimum was 36.6 and maximum was 67.4<sup>23</sup>.

**Statistical analyses.** Latent profile analysis (LPA) was used to make multiple-pollution profiles<sup>14</sup>. A series of LPAs was performed, ranging from two to seven latent profiles. The 12 air pollutants (PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub>, benzene, toluene, ethylbenzene, *m*-xylene, *p*-xylene, *o*-xylene and TBTEX) were used as LPA indicators. The grand mean centering of all 12 pollutants were used in the analysis to facilitate interpretation. Grand-mean centering subtracts the overall mean from a variable<sup>24</sup>. To identify the best fitting model, measures of relative statistical fit and the interpretability of their profile structure were used. Models with low values in AIC, BIC, aBIC metrics, and a significant Bootstrap Likelihood Ratio Test (BLRT) were preferred<sup>25</sup>. The absolute value of log likelihood is not recommended to be used for model selection, because this value gradually improves by adding more parameters to the model. Entropy was also evaluated for each model; and values closer to 1 suggest a higher discrimination of the latent classes<sup>26</sup>. In addition, the interpretability and parsimony of the candidate models were compared. Models with profiles including less than 5% of the class size are considered spurious<sup>27</sup>, and are not acceptable. Data preparation was done in Stata version 14. LPA was performed using Mplus version 7.4 (Muthen & Muthen, 1998–2015) mixture modeling procedure, with the robust maximum likelihood (MLR) estimator. Missing data were addressed using full information maximum likelihood (FIML). To examine how the various multiple-pollution profiles differ in terms of each air pollution component, one-way ANOVA and post hoc follow-up tests were used.

Kolmogorov Smirnov tests were used to test the normality of the pollutants and because the data was normally distributed, the Pearson's correlation test was used to estimate the correlation between pollutants. As the number of lung cancer cases was over-dispersed, negative binomial (NB) regression was performed to estimate the incidence rate ratios (IRR) and their 95% confidence intervals (CI) for each air pollutant and multiple pollution profiles, adjusted for age, sex, smoking at district level.

In order to calculate the population attributable fraction (PAF), the risk estimates for air pollutants were obtained from the results of negative binomial (NB) regression analysis.

The PAF for air pollutants as continuous variables, were estimated using the following equation<sup>28</sup>:

$$PAF = \frac{\exp[\text{Ln}(\text{RR}_{\text{unit}} \times \bar{x})] - 1}{\exp[\text{Ln}(\text{RR}_{\text{unit}} \times \bar{x})]}$$

In this equation, RR<sub>unit</sub> is the relative risk for each one unit increment in exposure to the air pollutant and  $\bar{x}$  is the average of exposure.

Statistical analyses were performed using Mplus version 7.4, Stata version 14 (Stata Corp LLC; College Station, TX, USA) and ArcGIS version 10.8.

As the data was inquired in aggregated form and anonymously, informed consent from individuals or their family was not required. This data is not publicly available, but can be inquired by formal request in aggregated and/or anonymous form from the Ministry of Health of Iran. The Ethic approval code of this project was IR.KMU.REC.1398.230. All methods were carried out in accordance with relevant guidelines and regulations.

## Results

Basic information about the area under study is shown in Table 1. The total number of lung cancer cases in 2014–2016 was 1850 in all districts of Tehran. We had to exclude the data of subjects who lived in remote suburbs of Tehran, which air pollutants estimation was not possible. Eventually, 1653 cases entered the final analyses. The distribution of lung cancer patients in different districts of Tehran is shown in Fig. 1. The highest number of patients per 100,000 populations was in regions 12, 6 and 11, respectively.

The spatial distributions and average levels of ambient air pollutants in different districts of Tehran are shown in Fig. 2. Districts with higher concentrations for pollutants were mostly in downtown (district 6, 7, 11, 12 and 14), and around the railway (district 16 and 17), and a few of the southern districts of the city (district 18). District 16 had the highest concentration of SO<sub>2</sub> and district 9, 2 and 6 had the highest concentration of NO, NO<sub>2</sub> and NO<sub>x</sub> pollutants during these years. District 12 had the highest concentration of VOC pollutants.

As indicated in Fig. 3, there was a strong correlation (Pearson's *r*) between pollutants, most notably for benzene compounds (benzene, toluene, ethylbenzene, *m*-xylene, *p*-xylene, *o*-xylene and TBTEX) (*p*-value < 0.001). The positive correlation between NO and NO<sub>x</sub> was weaker (*p*-value < 0.001).

Fit indices for the different LPA models are displayed in Table 2. All solutions provided acceptable classification accuracy, as indicated by entropy values close to 1. Although models with four, five, six, and seven latent class profiles had lower AIC, BIC, and aBIC than two and three latent class profiles, these models included classes with less than 1% of the sample. Therefore, the three latent profile model was preferred. The multi-pollution profiles are shown in Fig. 4. Profile 1 had the lowest scores for all pollutants, except SO<sub>2</sub>. We labeled this profile as "low multiple-pollution". Profile 3 had the highest scores of all pollutants. We labeled this profile as "high multiple-pollution". Profile 2 was in between and was labeled "medium multiple-pollution".

Summary statistics for each pollutant in different profiles are shown in Table 3. There was a significant difference between the means of all pollutants in the three profiles, except SO<sub>2</sub>.

Table 4 shows the IRR estimates by single-pollutant and multiple-pollutant multivariable negative binomial regression models, adjusted for age, gender, socioeconomic status, life expectancy and smoking prevalence. In single-pollutant models, *p*-xylene, *o*-xylene, ethylbenzene, benzene, *m*-xylene and TBTEX were significantly associated with increased lung cancer incidence in model 3, which was adjusted for age, gender, socioeconomic status, life expectancy and smoking prevalence.

In multi-pollutant models, the high multiple-air-pollutants profile was associated with higher lung cancer incidence when compared with the low multiple-air-pollutants profile.

The fraction of cancers attributable to air pollutant can be seen in Fig. 5. The highest fractions belong to *m*-xylene, *o*-xylene, and TBTEX.

<b>Number of districts</b>	<b>22</b>
Number of lung cancer patients	1653
<b>Age of patients (year)</b>	
Median (1st–3st quartile range)	65.5 (60–71)
<b>Gender of patients (male)</b>	
Percent (number)	70.1 (1158)
<b>Annual mean air pollutants</b>	
Median (1st–3st quartile range)	
PM <sub>10</sub> (µg/m <sup>3</sup> )	85.42 (65.5–99.9)
SO <sub>2</sub> (ppb)	50.82 (32.9–61.7)
NO <sub>2</sub> (ppb)	45.46 (41.6–50.2)
NO (ppb)	74.02 (55.6–98.5)
NO <sub>x</sub> (ppb)	112.21 (92.1–137.2)
Benzene (µg/m <sup>3</sup> )	9.05 (8.2–9.9)
Toluene (µg/m <sup>3</sup> )	26.35 (23.6–29.9)
Ethylbenzene (µg/m <sup>3</sup> )	6.49 (5.8–7.2)
<i>p</i> -xylene (µg/m <sup>3</sup> )	6.04 (5.6–6.6)
<i>o</i> -xylene (µg/m <sup>3</sup> )	6.60 (5.9–7.6)
<i>m</i> -xylene (µg/m <sup>3</sup> )	11.85 (10.4–13.4)
TBTEX (µg/m <sup>3</sup> )	64.68 (58.65–70.90)
<b>District level variables</b>	
Median (1st–3st quartile range)	
Total population (persons)	319,443 (263,536–421,225)
Life expectancy (year)	75.9 (75.4–76.7)
Population density (per km <sup>2</sup> )	158.5 (127.5–208.1)
Urban green space, per capita (m <sup>2</sup> per 1000 people)	13.9 (10.7–18.0)
Socioeconomic status*	49.95 (44.21–53.34)
Smoking prevalence in men (%)	15.7 (15.1–16.7)
Smoking prevalence in women (%)	0.79 (0.67–1.18)

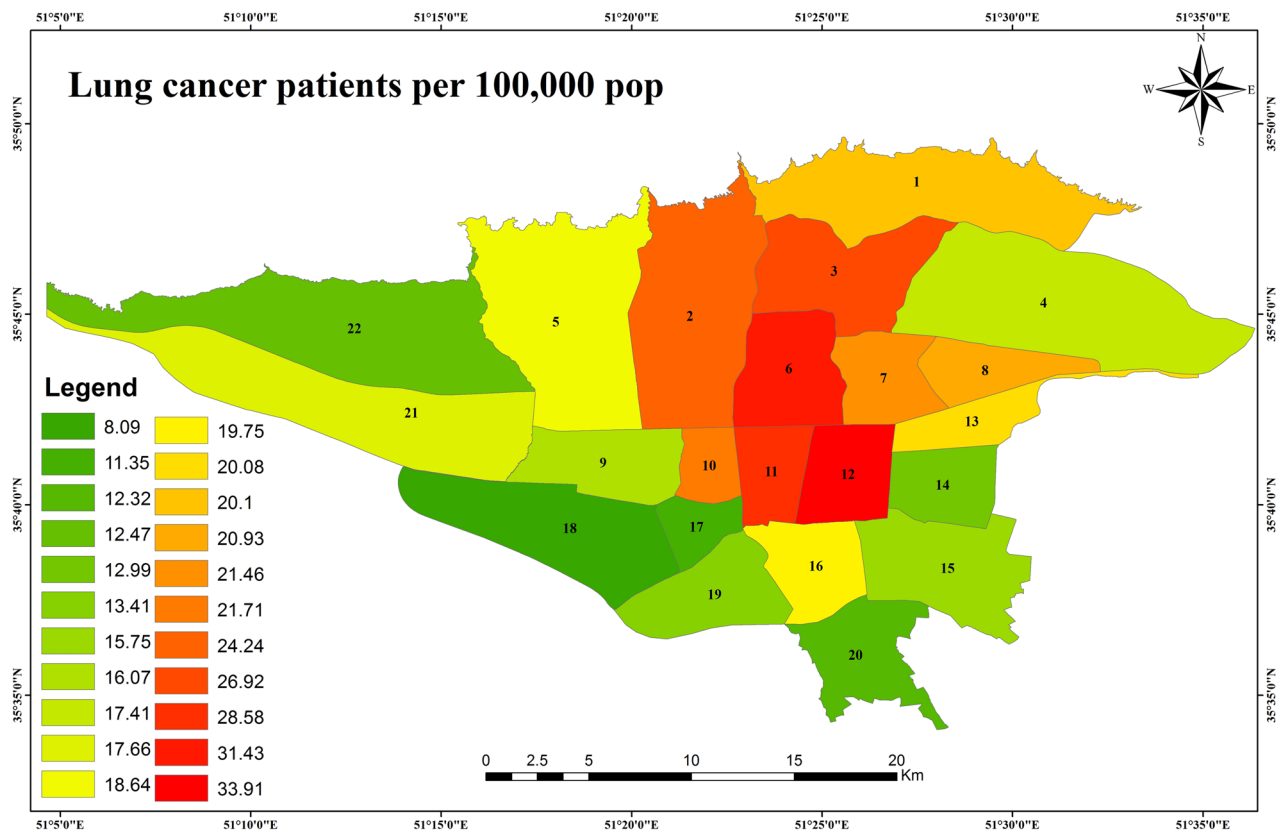
**Table 1.** Description of the study area, air pollution level, and district-level covariates. \*Socio-economic status score according to the 16 variables mentioned in the method section. This variable does not have a unit. The lowest value of this score was 36.6 and the highest was 67.4.

## Discussion

This study was the first to investigate the effect of single and multiple ambient air pollutants on lung cancer in Iran. The findings suggest that ambient air pollutants, especially *p*-xylene, *o*-xylene, ethylbenzene, benzene, *m*-xylene and TBTEX were associated with lung cancer. Previously several studies have also shown a strong association between air pollution and respiratory mortality<sup>29,30</sup> and respiratory diseases<sup>31,32</sup>, including chronic obstructive pulmonary disease (COPD), asthma, bronchitis, and decreased lung function<sup>33</sup>. Recently, some studies have shown the association between air pollution and lung cancer as well<sup>3</sup>.

Iran's national cancer registry data indicates an approximately sevenfold increase in the trend of lung cancer incidence over a 27-year span (1990 to 2016), in the whole country and in the capital city, Tehran<sup>34</sup>. In the past, lung cancer had been mainly attributed to direct tobacco smoke exposure. However, its increased incidence in never-smokers in the recent years shows that there are other risk factors that need to be discovered<sup>35</sup>. Some of the probable risk factors for lung cancer in never smokers could be environmental pollutants, such as air pollution, occupational carcinogens, radon and infections<sup>36</sup>. In Taiwan, air pollution was related with the incidence of lung cancer in never-smokers<sup>37</sup>; and the result of lung cancer screening programs in China and the United States in 2018 showed that the incidence of lung cancer in never smokers was significantly higher in China than the United States. Their data suggested that inclusion of ambient air pollution could improve the lung cancer risk models, especially for non-smokers<sup>38</sup>.

Air pollutants have been reported to be correlated in many studies. Faridi et al. reported positive correlations between PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub>, SO<sub>2</sub>, CO and O<sub>3</sub> in Tehran<sup>39</sup>, and another study from Los Angeles County, also reported correlations between multiple ambient air pollutants<sup>40,41</sup>. Studies from Spain have shown positive correlations between PM<sub>10</sub> and PM<sub>2.5</sub> and between nitrogen oxides (NO<sub>2</sub> and NO). The correlations between nitrogen oxides (NO and NO<sub>2</sub>) and particulate matter (PM<sub>10</sub>) is probably due to the common sources of these pollutants that are traffic, heating systems, industries, and other combustion processes<sup>42,43</sup>. High spatial correlations between exposure variables preclude the possibility to do multivariate adjusted analysis in air pollution studies. In the present study, because of the high correlation between pollutants, we used LPA models to investigate the association between multiple -pollutants and lung cancer incidence. Exposure profile modelling for multiple exposures has been used in previous epidemiologic studies on health outcomes such as blood pressure<sup>44</sup>, low birth weight<sup>40</sup>, total mortality<sup>45</sup>, respiratory mortality<sup>46</sup>, and lung cancer in nonsmokers<sup>47</sup>.



**Figure 1.** Spatial distribution of lung cancer patients (number of cases in 100,000) in different areas of Tehran in 2014–2016.

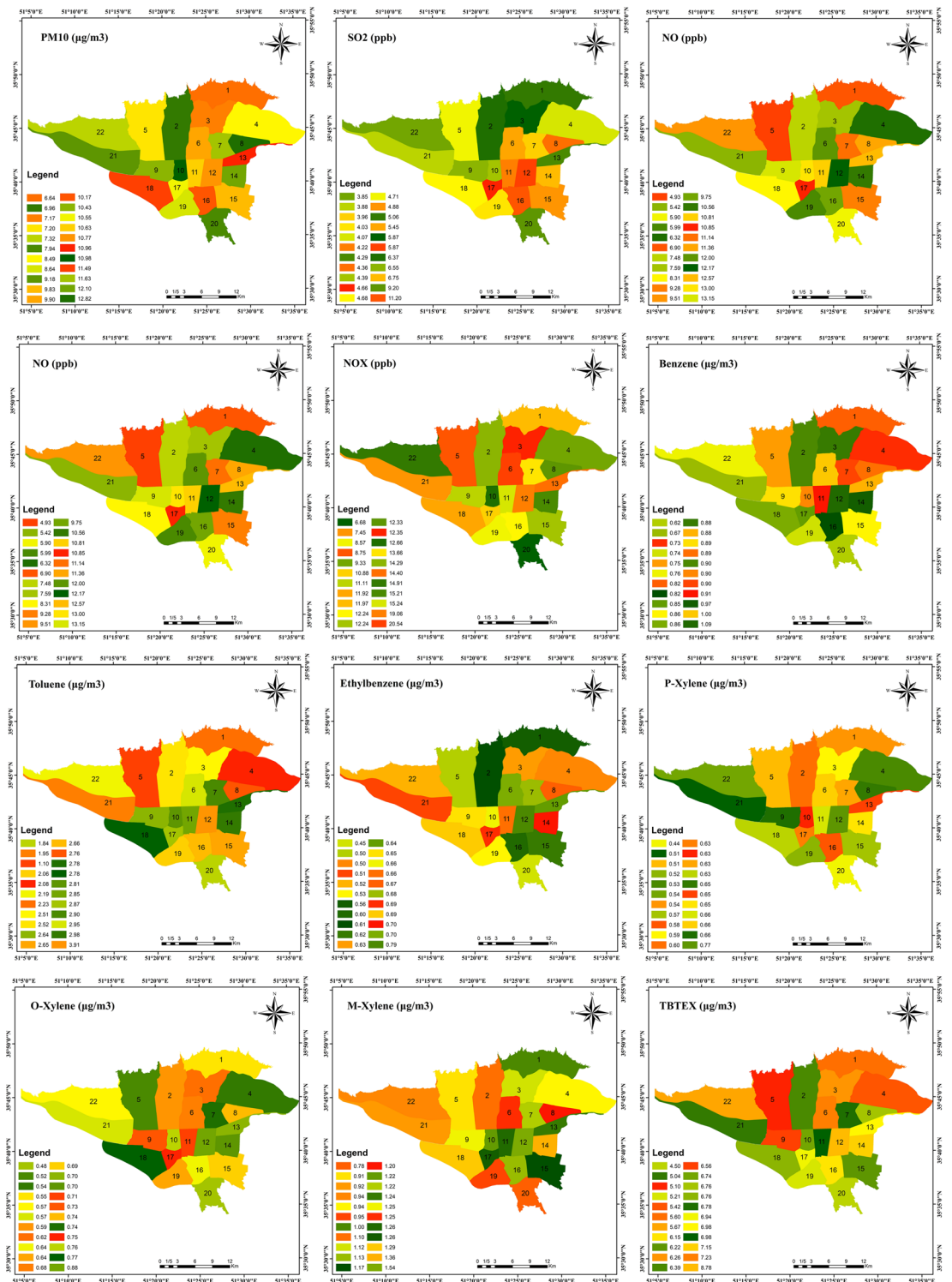
This method enables identifying possible unobservable subgroups, or latent classes, in a population using a number of related exposure variables and can help better understand the complex relations between risk factors and health outcomes, such as cancer, that may not be best explained by a single exposure<sup>13</sup>.

The concentration of BTEX in Tehran ambient air between 2005 and 2018 was estimated in a meta-analysis and risk assessment conducted by Abtahi et al. The rank of BTEX concentrations was benzene ( $149.18 \mu\text{g}/\text{m}^3$ : 31%) > *o*-xylene ( $127.16 \mu\text{g}/\text{m}^3$ : 27%) > ethylbenzene ( $110.15 \mu\text{g}/\text{m}^3$ : 23%) > toluene ( $87.97 \mu\text{g}/\text{m}^3$ : 19%). While in the present study toluene, *m*-xylene and benzene had the highest concentrations among VOCs<sup>20</sup>.

The primary sources of benzene and toluene in the ambient air of Tehran includes both mobile and stationary sources of emission. According to the results of Abtahi et al., the pooled concentrations of benzene ( $149.18 \mu\text{g}/\text{m}^3$ ) and *o*-xylene ( $125.57 \mu\text{g}/\text{m}^3$ ) in Tehran were higher than those in other regions around the world such as Ontario (Canada), Orleans (France), Bari (Italy), Kuala Lumpur (Malaysia) and Beijing; and the population residing in Tehran is at a considerable risk of exposure to carcinogens<sup>48</sup>. The inversion phenomenon, fossil fuel consumption of old vehicles, low-quality fuel, population congestion, and high-traffic highways, and the existence of several factories in the south of Tehran such as iron and steel industries, are other reasons for the high level of BTEX in Tehran city<sup>48,49</sup>.

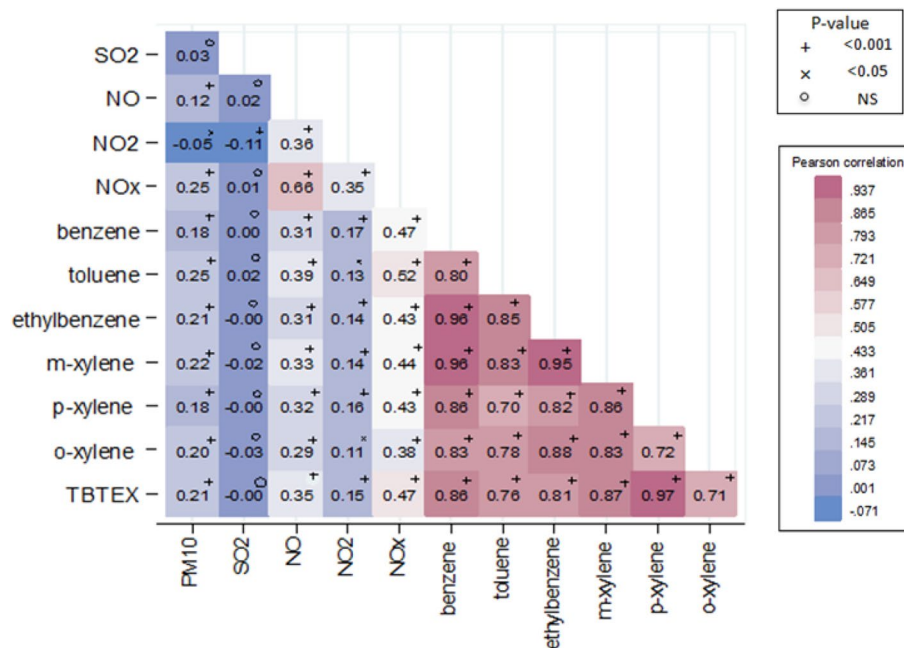
In the present study, about 70% of lung cancer patients were men and the prevalence of smoking among men was about 16%. But even after adjusting for smoking prevalence, the effect of air pollutants on lung cancer was significant.

Su et al. conducted an ecological study about ambient air pollution and all cancer incidences in Taiwan. Their results showed positive correlations between  $\text{PM}_{2.5}$ ,  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{O}_3$  levels and age-adjusted total cancer incidence rates<sup>50</sup>; and a study conducted on data from 2002 and 2011 in Brazil showed that traffic density and  $\text{NO}_2$  were associated with an increased incidence of respiratory cancers<sup>51</sup>. In a large population of 16,209 Norwegian men, after a 27-year follow-up the risk ratios for developing lung cancer attributed to  $\text{NO}_x$  and  $\text{SO}_2$  exposure were 1.08 (CI 95% = 1.06–1.11) and 1.03 (CI 95% = 0.77–1.38), respectively<sup>52</sup>. A cohort study conducted from January 1998 to December 2009 in four Northern Chinese cities including Tianjin, Shenyang, Taiyuan, and Rizhao, showed that the combined effect of  $\text{NO}_2$  and  $\text{PM}_{10}$  resulted in a significant increase in mortality from lung cancer<sup>53</sup>. Bai et al. in a population-based cohort study in Ontario, Canada (2001–2015) showed positive associations between lung cancer incidence with  $\text{PM}_{2.5}$  (hazard ratio [HR] = 1.02 [95% CI: 1.01–1.05] per  $5.3 \mu\text{g}/\text{m}^3$ ) and  $\text{NO}_2$  (HR = 1.05 [95% CI: 1.03–1.07] per 14 ppb), and each  $\sim 5 \mu\text{g}/\text{m}^3$  increase in outdoor  $\text{PM}_{2.5}$  concentration was associated with a 2% (95% CI: 1%–5%) increased risk of lung cancer. However, no associations were observed for  $\text{O}_3$  or  $\text{O}_x$  and lung cancer<sup>54</sup>. In our analysis,  $\text{NO}_2$  was associated with an increased risk of lung cancer as well. Also a meta-analysis of 14 studies showed that the pooled risk of  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  for lung cancer mortality was respectively RR: 1.14, CI95%: 1.07–1.21 and RR: 1.07, CI95%: 1.03–1.11<sup>55</sup>.



**Figure 2.** Spatial distribution and average levels of ambient air pollutants in different areas of Tehran in 2014–2016.

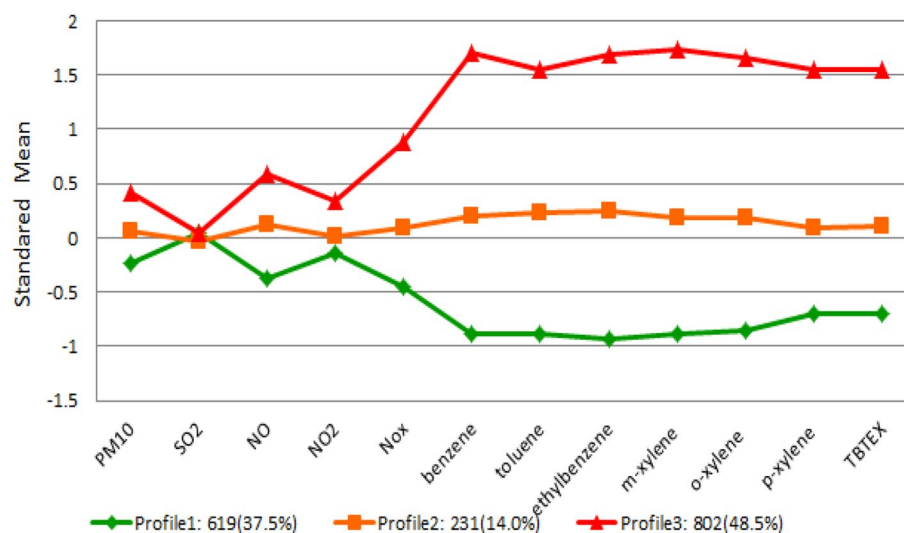
A population-based cohort study from the Korean National Health Insurance Service (NHIS) database on 2006–2007 data showed that they did not find an increased risk of lung cancer with higher exposure to PM<sub>10</sub> or NO<sub>2</sub>, in average concentrations of PM<sub>10</sub> = 60.9 µg/m<sup>3</sup> and NO<sub>2</sub> = 32.1 ppm<sup>56</sup>. The largest population-based case-control study of lung cancer among never-smoking females in Xuanwei and Fuyuan, China, done by Song et al. showed that a cluster of 25 PAHs had the strongest association with lung cancer [OR = 2.21; 95% CI = 1.67–2.87]; and nitrogen dioxide (NO<sub>2</sub>) was also directly related with lung cancer (OR = 2.06; 95% CI = 1.19, 3.49). However, neither benzo (a) pyrene (BaP) nor PM<sub>2.5</sub>, were associated with lung cancer in the



**Figure 3.** Pearson correlation matrix of air pollutants in the 22 districts of Tehran in 2014–2016.

	Log-likelihood	AIC	BIC	aBIC	Entropy	BLRT
2 Profile	-70,802.1	141,678.2	141,878.3	141,760.8	0.93	6922.7***
<b>3 Profile</b>	<b>-69,084.9</b>	<b>138,269.8</b>	<b>138,540.3</b>	<b>138,381.4</b>	<b>0.92</b>	<b>3434.3***</b>
4 Profile	-67,326.8	134,779.7	135,120.5	134,920.4	0.94	3516.1***
5 Profile	-66,087.4	132,326.8	132,738.0	132,496.5	0.96	2478.8***
6 Profile	-65,394.4	130,966.8	131,448.3	131,165.5	0.94	1386.1***
7 Profile	-64,790.5	129,785.1	130,336.9	130,012.8	0.95	1469.5***

**Table 2.** Fit indices for different models with number of profiles ranging from 2 to 7. *Note* AIC Akaike’s Information Criterion, BIC Bayesian Information Criterion, aBIC Sample Size Adjusted BIC, and BLRT bootstrap likelihood ratio test. \*\*\*P value <math><0.001</math>. Models with a low AIC, BIC, and aBIC value and high Entropy value as well as significant Bootstrap Likelihood Ratio Test (BLRT) were preferred. Bolded Fit Indices indicate that the 3 profile is optimal latent profile model, The selected (3 profile) has been bolded.



**Figure 4.** Standard mean values of pollutants in the three latent profiles in different areas of Tehran in 2014–2016.

Pollutant	Profile	Mean	SD	F	P-value
PM <sub>10</sub>	Profile 1	69.76	41.74	39.30	<0.001
	Profile 2	85.08	52.11		
	Profile 3	101.86	57.57		
SO <sub>2</sub>	Profile 1	52.42	37.78	1.22	0.295
	Profile 2	49.87	32.54		
	Profile 3	52.89	36.56		
NO	Profile 1	64.95	43.23	104.96	<0.001
	Profile 2	93.39	53.37		
	Profile 3	118.36	64.64		
NO <sub>2</sub>	Profile 1	46.65	16.42	19.22	<0.001
	Profile 2	49.46	16.32		
	Profile 3	54.78	21.23		
NO <sub>x</sub>	Profile 1	97.19	40.54	197.56	<0.001
	Profile 2	138.21	65.09		
	Profile 3	195.08	107.85		
Benzene	Profile 1	7.09	1.15	2132.15	<0.001
	Profile 2	9.79	1.10		
	Profile 3	13.57	2.15		
Toluene	Profile 1	20.76	3.63	1589.85	<0.001
	Profile 2	29.16	4.42		
	Profile 3	38.94	5.72		
Ethylbenzene	Profile 1	5.01	0.85	2664.12	<0.001
	Profile 2	7.07	0.76		
	Profile 3	9.60	1.06		
M-xylene	Profile 1	8.93	1.57	2371.73	<0.001
	Profile 2	13.01	1.63		
	Profile 3	18.87	3.21		
O-xylene	Profile 1	5.20	0.91	1694.88	<0.001
	Profile 2	7.42	1.28		
	Profile 3	10.59	1.69		
P-xylene	Profile 1	5.00	0.68	892.24	<0.001
	Profile 2	6.48	0.68		
	Profile 3	9.21	3.04		
TBTEX	Profile 1	51.76	7.89	938.71	<0.001
	Profile 2	69.73	8.38		
	Profile 3	102.18	35.34		

**Table 3.** The mean of air pollutants in different profiles.

multipollutant models<sup>57</sup>. Studies in Canada<sup>58,59</sup>, nine European countries<sup>60</sup>, the UK<sup>61</sup> and the US<sup>62</sup> also found that PM<sub>2.5</sub>, nitrogen oxides, nitrogen dioxide, and sulfur dioxide were associated with greater risk of lung cancer<sup>63</sup>. The ESCAPE study based in nine European countries concluded for every 10 µg/m<sup>3</sup> increase in PM<sub>10</sub>, the risk of lung cancer increases by 22%<sup>60</sup>.

**Strengths and limitation.** An important limitation of our study was the short time interval between the recorded pollutants and lung cancer incidence data. However, this was the latest available data about lung cancer in Tehran, at the time we started our study. Also, the study design was ecologic with no individual-level data, and the PAF would have been better estimated using individual-level data with adjustments for important confounding covariates. However, the results of this study still have important implications for public health, underscoring the need to reduce air pollution.

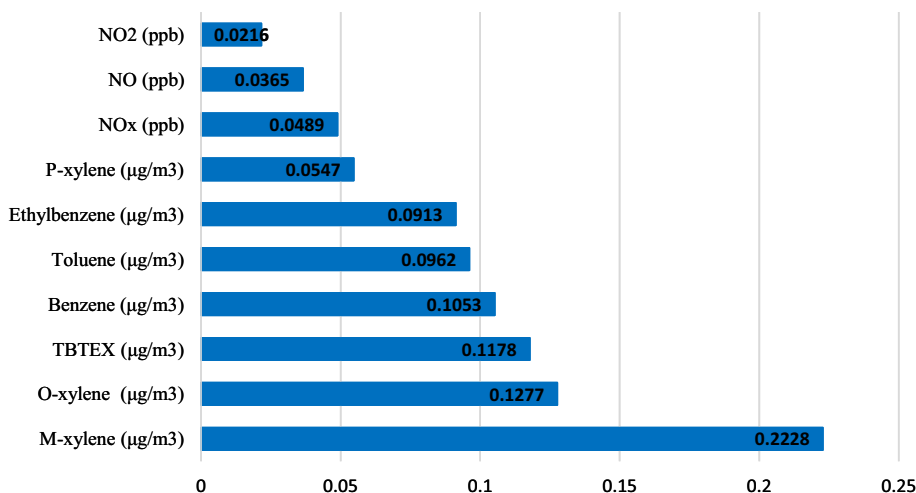
Tehran includes 22 districts, and the number of women with lung cancer was small in some districts, which prevented us to perform further analyses in gender subgroups. However, gender prevalence was adjusted in the analyses. Also, pathological data about cancer cell types was not available, and this prevented us to perform a separate analysis based on pathological type.

A novelty of our study was estimating the simultaneous effect of several air pollutants on lung cancer incidence. This helps to have a holistic picture of the effect of complex air pollution mixtures on human disease.



Pollutant	Model 1		Model 2		Model 3	
	RR (95% CI)	P-value	RR (95% CI)	P-value	RR (95% CI)	P-value
Single-pollutant						
PM <sub>10</sub> (µg/m <sup>3</sup> )	0.97 (0.89–1.05)	0.484	0.92 (0.84–1.002)	0.056	0.99 (0.86–1.14)	0.919
SO <sub>2</sub> (ppb)	0.93 (0.86–1.006)	0.074	0.93 (0.86–1.01)	0.124	0.98 (0.91–1.05)	0.591
NO <sub>2</sub> (ppb)	1.19 (1.04–1.37)	0.011	1.17 (1.008–1.36)	0.039	1.04 (0.91–1.19)	0.502
NO (ppb)	1.01 (0.96–1.06)	0.583	0.98 (0.92–1.05)	0.680	1.04 (0.99–1.09)	0.066
NO <sub>x</sub> (ppb)	1.04 (1.008–1.08)	0.016	1.04 (0.99–1.09)	0.064	1.05 (1.02–1.08)	0.001
Benzene (µg/m <sup>3</sup> )	2.15 (0.65–7.08)	0.206	1.50 (0.37–6.05)	0.567	3.86 (1.55–9.60)	0.004
Toluene (µg/m <sup>3</sup> )	1.14 (0.87–1.50)	0.326	1.03 (0.75–1.42)	0.813	1.50 (1.19–1.88)	<0.001
Ethylbenzene (µg/m <sup>3</sup> )	1.43 (0.29–6.88)	0.653	0.58 (0.08–3.93)	0.583	5.16 (1.23–21.57)	0.025
P-xylene (µg/m <sup>3</sup> )	2.79 (0.46–16.84)	0.261	1.29 (0.12–13.53)	0.827	9.41 (2.06–42.97)	0.004
O-xylene (µg/m <sup>3</sup> )	1.18 (0.29–4.83)	0.809	0.54 (0.10–2.74)	0.463	7.93 (1.74–36.21)	0.007
M-xylene (µg/m <sup>3</sup> )	1.23 (0.58–2.59)	0.580	0.83 (0.33–2.04)	0.690	2.63 (1.31–5.33)	0.007
TBTEX (µg/m <sup>3</sup> )	1.08 (0.94–1.24)	0.231	1.03 (0.87–1.22)	0.704	1.21 (1.08–1.34)	0.001
Multi-pollutant						
Profile 1 (low pollution)	Ref		Ref		Ref	
Profile 2 (moderate pollution)	1.006 (0.99–1.01)	0.165	1.009 (1.001–1.01)	0.019	1.001 (0.99–1.09)	0.897
Profile 3 (high pollution)	1.01 (1.001–1.01)	0.018	1.01 (1.009–1.02)	<0.001	1.01 (1.004–1.02)	0.002

**Table 4.** The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on lung cancer incidence in the districts of Tehran. Model 1: Adjusted for age and gender. Model 2: Adjusted for age, gender and urban green space, per capita (m<sup>2</sup> per 1000 people). Model 3: Adjusted for age, gender, socioeconomic status, life expectancy and smoking prevalence. The incidence rate ratio of lung cancer is estimated for each 10 unit increase in pollutants.



**Figure 5.** Estimated fraction of all lung cancer incidence attributable to each 1 unit increase in air pollutants in Tehran in 2014–2016.

## Conclusion

This is the first study to examine the associations between multiple air pollutants on lung cancer incidence in Iran. The findings suggest that lung cancer was associated with ambient air pollution in Tehran, and this association was stronger for p-xylene, o-xylene, ethylbenzene, benzene, m-xylene and TBTEX. Air pollution is a serious problem in Tehran, and decreasing the concentrations of air pollutants should be a key goal for policy makers to reduce the number of lung cancer cases in Tehran.

Received: 2 October 2020; Accepted: 15 April 2021

Published online: 29 April 2021

## References

1. WHO. *Cancer, Key facts*. <https://www.who.int/news-room/fact-sheets/detail/cancer>. Accessed 20 Jun 2020.

2. National Report of the National Cancer Registration Program for 2015. *Iran Ministry of Health and Medical Education. Non-communicable disease management center.* Cancer Department (2019).
3. Zhu, F. *et al.* 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 (2019).
4. Coker, E., Liverani, S., Su, J. G. & Molitor, J. Multi-pollutant modeling through examination of susceptible subpopulations using profile regression. *Curr. Environ. Health Rep.* **5**(1), 59–69 (2018).
5. Han, Y. *et al.* Association between size-segregated particles in ambient air and acute respiratory inflammation. *Sci. Total Environ.* **565**, 412–419 (2016).
6. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *IARC Monographs Volume 109: Outdoor Air Pollution. IARC Monogr Eval Carcinog Risks Hum, 109.* International Agency for Research on Cancer, Lyon 9–444. <https://monographs.iarc.fr/iarc-monographs-on-the-evaluationof-carcinogenic-risks-to-humans-7/>. (2016).
7. Hvidtfeldt, U. *et al.* Low-level ambient air pollution exposure and risk of lung cancer: A pooled analysis of 7 European cohorts. *Environ. Epidemiol.* **3**, 171 (2019).
8. Dominici, F., Peng, R. D., Barr, C. D. & Bell, M. L. Protecting human health from air pollution: Shifting from a single-pollutant to a multi-pollutant approach. *Epidemiology* **21**(2), 187 (2010).
9. Taylor, K. W. *et al.* Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: Lessons from an innovative workshop. *Environ. Health Perspect.* **124**(12), A227–A229 (2016).
10. Stafoggia, M., Breitner, S., Hampel, R. & Basagaña, X. Statistical approaches to address multi-pollutant mixtures and multiple exposures: The state of the science. *Curr. Environ. Health Rep.* **4**(4), 481–490 (2017).
11. Davalos, A. D., Luben, T. J., Herring, A. H. & Sacks, J. D. Current approaches used in epidemiologic studies to examine short-term multipollutant air pollution exposures. *Ann. Epidemiol.* **27**(2), 145–153 (2017).
12. Caban-Martinez, A. J., Santiago, K. M., Baum, J. & Solle, N. N. S. Silicone-based wristband passive samplers in the detection of firefighter occupational carcinogenic exposures. *Cancer Prev. Res.* **13**, 45 (2020).
13. Kolpacoff, V. L. *Latent Class Modeling of Environmental Exposures for North Carolina at the Census Tract Level* (2019).
14. Hagenars, J. A. & McCutcheon, A. L. *Applied Latent Class Analysis* (Cambridge University Press, Cambridge, 2002).
15. Guo, H., Chang, Z., Wu, J. & Li, W. Air pollution and lung cancer incidence in China: Who are faced with a greater effect?. *Environ. Int.* **132**, 105077 (2019).
16. Fuller, C. H., Feeser, K. R., Sarnat, J. A. & O'Neill, M. S. Air pollution, cardiovascular endpoints and susceptibility by stress and material resources: A systematic review of the evidence. *Environ. Health* **16**(1), 58 (2017).
17. UNFPA Iran: *Population study and census data from the Statistical Center of Iran.* <https://iran.unfpa.org/Documents/Census2011/2011%20Census%20Selected%20Results%20-%20Eng.pdf>. Accessed 20 Jun 2020.
18. Amini, H. *et al.* Land use regression models to estimate the annual and seasonal spatial variability of sulfur dioxide and particulate matter in Tehran, Iran. *Sci. Total Environ.* **488**, 343–353 (2014).
19. Amini, H. *et al.* Annual and seasonal spatial models for nitrogen oxides in Tehran, Iran. *Sci. Rep.* **6**(1), 1–11 (2016).
20. Amini, H. *et al.* Spatiotemporal description of BTEX volatile organic compounds in a Middle Eastern megacity: Tehran study of exposure prediction for environmental health research (Tehran SEPEHR). *Environ. Pollut.* **226**, 219–229 (2017).
21. Amini, H., Schindler, C., Hosseini, V., Yunesian, M. & Künzli, N. Land use regression models for Alkylbenzenes in a middle eastern megacity: Tehran study of exposure prediction for environmental Health Research (Tehran SEPEHR). *Environ. Sci. Technol.* **51**(15), 8481–8490 (2017).
22. Asadi-Lari, M. *et al.* Response-oriented measuring inequalities in Tehran: Second round of UrbanHealth Equity Assessment and Response Tool (Urban HEART-2), concepts and framework. *Med. J. Islam Repub. Iran* **27**(4), 236 (2013).
23. Sadeghi, R. & Zanjari, N. The inequality of development in the 22 districts of Tehran metropolis. *Soc. Welf. Q.* **17**(66), 149–184 (2017).
24. Muthén, M. (1998–2015) *Mplus User's Guide* (Muthen & Muthen, Los Angeles, 2015).
25. Nylund, K. L., Asparouhov, T. & Muthén, B. O. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Struct. Equ. Model.* **14**(4), 535–569 (2007).
26. Muthén, L. & Muthén, B. Re: *What is a Good Value of Entropy.* [Online comment]. <http://wwwstatmodel.com/discussion/messages/13/2562.html>. (2017).
27. Marsh, H. W., Lüdtke, O., Trautwein, U. & Morin, A. J. Classical latent profile analysis of academic self-concept dimensions: Synergy of person-and variable-centered approaches to theoretical models of self-concept. *Struct. Equ. Modeling* **16**(2), 191–225 (2009).
28. Azevedo e Silva, G. *et al.* The fraction of cancer attributable to ways of life, infections, occupation, and environmental agents in Brazil in 2020. *PLoS ONE* **11**(2), e0148761 (2016).
29. Dastoorpoor, M. *et al.* Short-term effects of air pollution on respiratory mortality in Ahvaz, Iran. *Med. J. Islam. Repub. Iran* **32**, 30 (2018).
30. Dehghan, A., Khanjani, N., Bahrampour, A., Goudarzi, G. & Yunesian, M. The relation between air pollution and respiratory deaths in Tehran, Iran-using generalized additive models. *BMC Pulm. Med.* **18**(1), 49 (2018).
31. Dastoorpoor, M. *et al.* Associations of short-term exposure to air pollution with respiratory hospital admissions in Ahvaz, Iran. *Process Saf. Environ. Prot.* **123**, 150–160 (2019).
32. Vahedian, M., Khanjani, N., Mirzaee, M. & Koolivand, A. Associations of short-term exposure to air pollution with respiratory hospital admissions in Arak, Iran. *J. Environ. Health Sci. Eng.* **15**(1), 17 (2017).
33. Raji, H. *et al.* Acute effects of air pollution on hospital admissions for asthma, COPD, and bronchiectasis in Ahvaz, Iran. *Int. J. Chronic Obstr. Pulm. Dis.* **15**, 501 (2020).
34. Rajai, N. *et al.* National and subnational trends in incidence and mortality of lung cancer in Iran from 1990 to 2016. *Asia-Pac. J. Clin. Oncol.* **16**(129), 136 (2020).
35. Chen, Y.-J. *et al.* Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression. *Cell* **182**(1), 226–244 (2020).
36. Rait, G. & Horsfall, L. Twenty-year sociodemographic trends in lung cancer in non-smokers: A UK-based cohort study of 3.7 million people. *Cancer Epidemiol.* **67**, 101771 (2020).
37. Tseng, C.-H. *et al.* The relationship between air pollution and lung cancer in nonsmokers in Taiwan. *J. Thorac. Oncol.* **14**(5), 784–792 (2019).
38. Yang, D., Liu, Y., Bai, C., Wang, X. & Powell, C. A. Epidemiology of lung cancer and lung cancer screening programs in China and the United States. *Cancer Lett.* **468**, 82–87 (2020).
39. Faridi, S. *et al.* Spatial homogeneity and heterogeneity of ambient air pollutants in Tehran. *Sci. Total Environ.* **697**, 134123 (2019).
40. Coker, E. *et al.* Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. *Environ. Int.* **91**, 1–13 (2016).
41. Tavallali, P., Gharibi, H., Singhal, M., Schweizer, D. & Cisneros, R. A multi-pollutant model: A method suitable for studying complex relationships in environmental epidemiology. *Air Qual. Atmos. Health* **13**, 645–657 (2020).
42. Rovira, J., Domingo, J. L. & Schuhmacher, M. Air quality, health impacts and burden of disease due to air pollution (PM10, PM2.5, NO2 and O3): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain). *Sci. Total Environ.* **703**, 135538 (2020).

43. Núñez-Alonso, D., Pérez-Arribas, L. V., Manzoor, S. & Cáceres, J. O. Statistical tools for air pollution assessment: Multivariate and spatial analysis studies in the Madrid region. *J. Anal. Methods Chem.* <https://doi.org/10.1155/2019/9753927> (2019).
44. Molitor, J. *et al.* Blood pressure differences associated with optimal macronutrient intake trial for heart health (OMNIHEART)-like diet compared with a typical American diet. *Hypertension* **64**, 1198–1204. <https://doi.org/10.1161/HYPERTENSIONAHA.114.03799> (2014).
45. Zanobetti, A., Austin, E., Coull, B. A., Schwartz, J. & Koutrakis, P. Health effects of multi-pollutant profiles. *Environ. Int.* **71**, 13–19 (2014).
46. Pirani, M. *et al.* Analysing the health effects of simultaneous exposure to physical and chemical properties of airborne particles. *Environ. Int.* **79**, 56–64 (2015).
47. Papatomas, M., Molitor, J., Richardson, S., Riboli, E. & Vineis, P. Examining the joint effect of multiple risk factors using exposure risk profiles: Lung cancer in nonsmokers. *Environ. Health Perspect.* **119**(1), 84–91 (2011).
48. Abtahi, M. *et al.* The concentration of BTEX in the air of Tehran: A systematic review-meta analysis and risk assessment. *Int. J. Environ. Res. Public Health* **15**(9), 1837 (2018).
49. Dehghani, M. H., Sanaei, D., Nabizadeh, R., Nazmara, S. & Kumar, P. Source apportionment of BTEX compounds in Tehran, Iran using UNMIX receptor model. *Air Qual. Atmos. Health* **10**(2), 225–234 (2017).
50. Su, S.-Y. *et al.* Associations between ambient air pollution and cancer incidence in Taiwan: An ecological study of geographical variations. *BMC Public Health* **19**(1), 1–8 (2019).
51. Ribeiro, A. G. *et al.* Incidence and mortality for respiratory cancer and traffic-related air pollution in São Paulo, Brazil. *Environ. Res.* **170**, 243–251 (2019).
52. Nafstad, P. *et al.* Urban air pollution and mortality in a cohort of Norwegian men. *Environ. Health Perspect.* **112**(5), 610–615 (2004).
53. Chen, X. *et al.* Long-term exposure to urban air pollution and lung cancer mortality: A 12-year cohort study in Northern China. *Sci. Total Environ.* **571**, 855–861 (2016).
54. Bai, L. *et al.* Exposure to ambient air pollution and the incidence of lung cancer and breast cancer in the Ontario Population Health and Environment Cohort. *Int. J. Cancer* **146**(9), 2450–2459 (2020).
55. Kim, H.-B., Shim, J.-Y., Park, B. & Lee, Y.-J. Long-term exposure to air pollutants and cancer mortality: A meta-analysis of cohort studies. *Int. J. Environ. Res. Public Health* **15**(11), 2608 (2018).
56. Moon, D. H., Kwon, S. O., Kim, S.-Y. & Kim, W. J. Air pollution and incidence of lung cancer by histological type in Korean adults: A Korean National Health Insurance Service Health Examinee Cohort Study. *Int. J. Environ. Res. Public Health* **17**(3), 915 (2020).
57. Vermeulen, R. *et al.* Constituents of household air pollution and risk of lung cancer among never-smoking women in Xuanwei and Fuyuan, China. *Environ. Health Perspect.* **127**(9), 097001 (2019).
58. Weichenthal, S. *et al.* Long-term exposure to ambient ultrafine particles and respiratory disease incidence in Toronto, Canada: a cohort study. *Environ. Health* **16**(1), 64 (2017).
59. Hystad, P., Demers, P. A., Johnson, K. C., Carpiano, R. M. & Brauer, M. Long-term residential exposure to air pollution and lung cancer risk. *Epidemiology* **24**, 762–772 (2013).
60. Raaschou-Nielsen, O. *et al.* Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol.* **14**(9), 813–822 (2013).
61. Atkinson, R. *et al.* Long-term exposure to outdoor air pollution and the incidence of chronic obstructive pulmonary disease in a national English cohort. *Occup. Environ. Med.* **72**(1), 42–48 (2015).
62. Pope, C. A. III. *et al.* Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: Shape of the exposure–response relationships. *Environ. Health Perspect.* **119**(11), 1616–1621 (2011).
63. Hamra, G. B. *et al.* Lung cancer and exposure to nitrogen dioxide and traffic: A systematic review and meta-analysis. *Environ. Health Perspect.* **123**(11), 1107–1112 (2015).

## Acknowledgements

This study was financially supported by Kerman University of Medical Sciences by Grant No. 98000297. The authors thank the Cancer Department of the Ministry of Health of Iran, and Tehran Air Quality Control Company (AQCC), which provided data for this study.

## Author contributions

Study design: Z.K., N.K. and M.R. Conceptualization: Z.K., N.K., M.P. and M.R. Data cleaning: Z.K. and M.P. Data management: K.E., H.A., S.M.T.-S., N.K. Statistical analysis and Statistical inference: Z.K., M.R., M.P. and H.A. Writing the original draft: Z.K., N.K. and M.R. Review and editing: Z.K., N.K., M.P., M.R., K.E., S.M.T.-S., N.K., H.A. All authors participated in revision of the final draft and agreed on the final content.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to N.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021