BMJ Open

Multiple Air Pollutants Exposure and Leukemia Incidence in Tehran, Iran

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-060562
Article Type:	Original research
Date Submitted by the Author:	30-Dec-2021
Complete List of Authors:	khorrami, zahra; Kerman University of Medical Sciences, Epidemiology & Biostatistics Pourkhosravani, Mohsen; Shahid Bahonar University of Kerman Eslahi, Marzieh; Kerman University of Medical Sciences, Department of Epidemiology and Biostatistics Rezapour, Maysam; Mazandaran University of Medical Sciences Akbari, Mohammad Esmail; Shahid Beheshti University of Medical Sciences Amini, Heresh; University of Copenhagen Taghavi-Shahri, Seyed Mahmood; University of Copenhagen, Department of Public Health Künzli, Nino ; Schweizerisches Tropen- und Public Health-Institut Etemad, Koorosh; Shahid Beheshti University of Medical Sciences, Department of Epidemiology Khanjani, Narges ; Shahid Beheshti University of Medical Sciences
Keywords:	Leukaemia < HAEMATOLOGY, EPIDEMIOLOGY, Epidemiology < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Multiple Air Pollutants Exposure and Leukemia Incidence in Tehran, Iran

Zahra Khorrami[1], Mohsen Pourkhosravani[2], Marzieh Eslahi[3], Maysam Rezapour[4], Mohammad Esmail Akbari[5], Heresh Amini[6], Seyed Mahmood Taghavi-Shahri[7], Nino Künzli[8, 9], Koorosh Etemad[10*], Narges Khanjani[11, 12*]

1. Ophthalmic Research Center, Research Institute for Ophthalmology and Vision Science, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Department of Geography and Urban Planning, Shahid Bahonar University of Kerman, Kerman, Iran
3. Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran
4. Amol Faculty of Paramedical Sciences, Mazandaran University of Medical Sciences, Sari, Iran.
5. Cancer Research Center (CRC), Shahid Beheshti University of Medical Sciences, Tehran, Iran.
6. Department of Public Health, University of Copenhagen, Copenhagen, Denmark
7. Section of Environmental Health, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
8. Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland
9. University of Basel, Basel, Switzerland
10. Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran
11. Environmental Health Engineering Research Center, Kerman University of Medical Sciences, Kerman, Iran
12. Monash Centre for Occupational & Environmental Health, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

*Corresponding authors:
1. Prof. Narges Khanjani, Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran.
Tel/Fax: 034-3132-5102 Email: n_khanjani@kmu.ac.ir.
2. Koorosh Etemad, Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
Tel/Fax: 021- 2243-2040 Email: etemadk@gmail.com

Abstract

Objective: Leukemia is one of the most common cancers and may be associated with exposure to environmental carcinogens, especially outdoor air pollutants. The objective of this study was to investigate the association of air pollution and leukemia in Tehran, Iran.

Design: Data about the residential location of leukemia cases diagnosed from 2010 to 2016 were inquired from the Ministry of Health cancer database. Data from a previous study was used to determine long-term average exposure to different air pollutants in 22 districts of Tehran. Latent profile analysis (LPA) was used to classify pollutants in two exposure profiles. The association between air pollutants and leukemia incidence was analyzed by negative binomial regression.

Setting: Twenty-two districts of Tehran megacity.

Participants: Leukemia patients.

Outcome measures: The outcome variables were incidence rate ratios Myeloid and Lymphoid Leukemia across the districts of Tehran.

Results:

The districts with higher concentrations for all pollutants were near city center. The incidence rate ratio (IRR) was positive but non-significant for most of the air pollutants. However, annual mean NOx was directly and significantly associated with leukemia incidence in the fully adjusted model (IRR (95% CI): 1.03 (1.003, 1.06) per 10 ppb increase). Based on LPA, districts with a higher multiple air-pollutants profile were also associated with higher leukemia incidence (IRR (95% CI): 1.003 (0.99, 1.007) per 1 ppb increase).

Conclusions:

Our study shows that districts with higher air pollution have higher incidence rate of leukemia cancer in Tehran, Iran. This study warrants conducting further research with individual data and better control of confounding.

Keywords

Air pollution ; multiple pollutants; pollutant profile; Leukemia; Latent profile analysis

Strengths and limitations of this study

- The first study for estimating the simultaneous effect of single and multiple ambient air pollutants on the incidence of leukemia in Iran.
- We included potential confounding covariates at area level based on administrative survey data
- Despite this, control for confounding variables confounders are not done at the individual level.
- Also, the associations in this study ignored within-area variation.

Background

Leukemia is one of the most common hematological malignancies among adults and children [1]. In 2020, leukemia was amongst the 15 more common cancers worldwide and contributed to 2.5 % of all new cases of cancer, and 3.1% of all cancer death. Worldwide, the age standardized incidence and mortality rates of leukemia are 6.3 per 100,000 and 4.0 per 100,000 for men and 4.5 per 100,000 and 2.7 per 100,000 in women, respectively. In countries with low and medium human development index (HDI), the age- standardized incidence and mortality rates of leukemia are, respectively, 3.8 per 100,000 and 3.0 per 100,000 for men, and 2.9 per 100,000 and 2.0 per 100,000 for women [2]. In Iran where cancer is the third leading cause of death after cardiovascular disease mortality and traffic accidents, leukemia is the 6th most common cancer in men with age standardized rates of 8.32 per 100,000, and the 5th in women with age standardized rates of 5.40 per 100,000 [3].

Outdoor air pollution may have serious adverse effects on human health, including cancer [4]. It has been proven that some environmental factors including ambient air pollution are carcinogenic to humans [5].

The etiology of leukemia remains mostly unknown. Some studies have reported that adverse gene-environment interactions may lead to leukemia [1]. A recent meta-analysis on the association between outdoor air pollution and childhood leukemia showed a positive linear association between benzene exposure and risk of childhood leukemia, particularly for acute myeloid leukemia, among children under 6 years of age, and nitrogen dioxide at high levels was also associated with leukemia [6].

In the current study, we aimed to examine the association between estimates of district level averages of multiple ambient air pollutants and leukemia incidence across Tehran districts.

Methodology

Research location

The present study was carried out based on annual mean air pollution levels in 22 districts of Tehran megacity, which is the capital of Iran. According to the World Population Review report, Tehran's 2021 population is now estimated to be 9.2 million inhabitants. Tehran includes 22

districts in which each include from 174,239 to 919,001 residents according to the latest 2016 census[7]. Population density is higher in the central, western, and southern regions[8]. Tehran suffers from severe air pollution as documented by numerous studies [9-12]. The central districts with higher population densities (five districts of 2, 6, 10, 11, and 12 at the center of Tehran) encounter more air pollution [8].

Data Sources

Leukemia data

Information about the leukemia patients (acute lymphoblastic leukemia (ALL) or acute myeloid leukemia (AML) residing in Tehran, diagnosed between 2010 and 2016 alongside their residential address were obtained from the Ministry of Health's Cancer Registry, on a district basis.

Exposure assessment

The annual mean concentrations of several pollutants in the 22 districts of Tehran were obtained from land use regression (LUR) models developed in previous studies, for PM_{10}, SO_2, NO, NO_2, and NO_X in Tehran, which were based on measurements conducted at 23 regulatory network monitoring sites in 2010 [10, 11]. Data about volatile organic compounds (VOCs) were taken from a previous study [12]. The average of sampling site estimates for each pollutant, in each district was determined and included in the analyzes of this study.

Covariates

District level data including urban green space per capita, life expectancy, and socioeconomic status, was extracted from the Urban Health Equity Assessment and Response Tool (Urban HEART-2), which has been conducted in the 22 districts of Tehran. A detailed description of the Urban HEART-2 study can be found elsewhere **[13]**

The socio-economic indicators of the 22 districts of Tehran was extracted from a study conducted by Sadeghi et al [14].

Statistical analyses

This study used latent profile analysis (LPA) to investigate the patterns of multiple air pollutants exposure. LPA is a statistical method to identify unobserved subgroups (profile) within

populations based on observed variables. LPA has several advantages over traditional methods, such as cluster analysis, e.g. it does not require researchers to specify the number of profiles in advance, which is more suitable for addressing research questions that are exploratory in nature. Moreover, unlike cluster analysis that assigns individuals to clusters absolutely, LPA assigns individuals to subgroups probabilistically[15].

Most air pollutants in this study had a skewed distribution and were transformed by natural logarithm before LPA, except PM_{10} and SO_2. Several latent profile models were performed, ranging from two to five latent profiles. The most appropriate number of subgroups was identified based on statistical criteria and profile interpretability. The statistical criteria included the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample-size adjusted Bayesian information criterion (a-BIC), the Lo-Mendell-Rubin likelihood ratio test (LMR-LRT), and the Vuong-Lo-Mendell-Rubin likelihood ratio tests (VLMR-LRT). Smaller values of the AIC, BIC, and aBIC indicate a better fit model. A significant p-value of LMR-LRT and VLMR-LRT indicates that the k class model is preferred over the k-1 class model [16].

Data preparation was done in Stata version 14. LPA was conducted using Mplus version 7.4. One-way ANOVA and post hoc follow-up tests were used to investigate the differences between profiles of multiple pollutions in terms of each component of air pollution. Management of missing data and other statistical preparation details have been mentioned in our previous publication **[7]**.

Kolmogorov Smirnov test was used to test the normality of the pollutants data and because the data were not normally distributed, Spearman's correlation test was used to estimate the correlation between pollutants. As the number of leukemia 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, and district level (by using them as a covariate). Statistical analyses were performed using Mplus version 7.4, Stata version 14 (Stata Corp LLC; College Station, TX, USA) and ArcGIS version 10.8 was used for exposure assignment to districts and data visualization.

Patient and public involvement

Patients and the public were not involved in the design and conduct of this research.

Results

Summary statistics of patients, air pollutants, and information about the area under study are shown in Table 1. The total number of leukemia cases in 2010–2016 in all districts of Tehran was 3,237. The distribution of leukemia patients in different districts of Tehran is shown in Figure 1. The highest numbers of leukemia patients per 100,000 people were in districts 6, 3, 7 and 2, respectively (Figure 1).

As shown in Figure 2, there was a strong correlation between pollutants, most notably for VOCs (benzene, toluene, m-xylene, p-xylene, o-xylene and TBTEX) (p-value<0.001). The positive correlation between NO and NOx was weaker (r=0.56).

Fit indices for the different LPA models are displayed in Table 2. Several latent profile models were considered for selecting the best model. Although, the AIC, BIC, and aBIC of the 2-profile were more than other models, the LMR of LRT (Lo–Mendell–Rubin likelihood ratio test) and VLMR (Vuong-Lo-Mendell-Rubin likelihood ratio test) were significant and interpretable in the two-profile model.

Figure 3 shows the multi-pollution profiles. Profile 1 had the lowest scores for all pollutants including (PM_{10}, SO2, NO, NO_2, NO_X, benzene, toluene, ethylbenzene, m-xylene, p-xylene, o-xylene and TBTEX). We labeled this profile as “low 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 two profiles, except for SO_2.

Table 4 shows the IRR estimates and 95% confidence interval (CI) by single-pollutant and multiple-pollutant multivariable negative binomial regression models, adjusted for age, gender, socioeconomic status and life expectancy.

In single-pollutant models, NO_2 and NO_X were significantly associated with increased leukemia incidence in model 1 (adjusted for age and gender), and also in model 2 (adjusted in addition for urban green space per capita) with IRR (95% CI) of model 2, respectively, 1.35 (1.11-1.64) and 1.07 (1.03-1.11) both per 10 ppb increase in pollutants. In model 3, after adjustment for age, gender, socioeconomic status, and life expectancy, only NO_X was significantly associated with increased leukemia incidence with IRR (95% CI) of 1.03 (1.003-1.06) per 10 ppb increase in

NOx. However, NO_2 was borderline significantly associated with increased leukemia incidence (IRR=1.19, CI 95%=0.99-1.43) in this model.

In multi-pollutant models, the high multiple-air-pollutants profile was associated with higher leukemia incidence when compared with the low multiple-air-pollutants profile, but this association did not reach significance.

Tables 5 and 6 show the IRR (95% CI) estimates by single-pollutant and multiple-pollutant multivariable negative binomial regression models for myeloid and lymphoid leukemia incidence, respectively. NO, NO_2 and NO_x were related to increased lymphoid leukemia incidence, while NO_2 and TBTEX were related to increased myeloid leukemia incidence.

Discussion

The findings of our study show that exposure to air pollutants, such as nitrogen oxides and VOCs may be associated with the incidence of leukemia. Our study was the first to investigate the effect of single and multiple ambient air pollutants on leukemia in Iran.

Leukemia is one of the most common cancers in children and adults. The cause of leukemia is currently unknown [17]. However, some sources have suggested that genetic and transgenic mutations due to environmental factors may contribute to leukemia [1, 18]. A study in Shanghai, China showed that air pollution from industrial waste gas emissions was associated with the incidence of several cancers including leukemia [19].

Many experimental studies have shown that air pollutants may be carcinogenic by causing DNA damage and mutations[20, 21], [22] [23] [24, 25].

A review indicated that exposure to air pollutants is associated stronger with leukemia, than other cancers [17]. Carlos-Wallace et al. also conducted a meta-analysis and reported associations between childhood leukemia and benzene exposure. They indicated that in studies that evaluated benzene versus all solvents, the pooled RR (Relative Risk from meta-analysis) was higher for maternal exposure compared to paternal exposure. For NO_2, an excess risk was reported in concentration-response meta-analysis from 40 $\mu g/m^3$ to 60 $\mu g/m^3$; however, the increase was not statistically significant and was mainly related to ALL [6].

A study in the United Kingdom revealed that there was an increased risk of leukemia from low-level exposure to benzene from smoking, and that benzene may contribute to up to a third of smoking-induced leukemia. The contribution of benzene to smoking-induced leukemia was estimated to be between 9% and 24%. For AML this contribution was estimated as 11-30% [26]. Janitz et al. conducted a case-control study between 1997 and 2012 including 307 patients with acute leukemia from the Oklahoma Central Cancer Registry and 1013 controls matched on week of birth from birth certificates and showed no association between benzene and childhood leukemia, in Oklahoma. However, based on the analysis of quartiles they indicated that there was a positive concentration-response relation between benzene and AML [27].

Findings of the population-based study by Ribeiro et al in São Paulo, Brazil has shown that NO_2 and traffic density were associated with Hodgkin lymphoma and lymphoid leukemia in children; and the variations in the incidence rate ratios in gender and SES groups may be because of differences in underlying risk and exposure profiles [28]. Raaschou-Nielsen et al. conducted a nationwide case-control study in Denmark and indicated that long-term exposure to traffic-related air pollutants (NO_x and NO_2) was associated with acute myeloid leukemia, but not other subtypes of leukemia, in the general population [29]. A Canadian population-based case–control study revealed a weak association between all forms of leukemia only at low concentrations of NO_2. The study showed an 'n-shaped' response function between exposure to NO_2 and all forms of leukemia. The OR was 1.20 (95% CI: 0.97-1.48) from the 10^{th} percentile (4.51 ppb) to the median (14.66 ppb), then the OR decreased to 0.79 (95% CI: 0.68, 0.93) from the 75th percentile (22.75) to the 90th (29.7 ppb) [30].

In a large nationwide population-based case-control study in Denmark, Taj et al. showed a high risk for overall leukemia in association with secondary inorganic aerosols including nitrate (NO_3), and that AML was associated with NO_3 [31].

Studies focusing on the association between childhood leukemia and exposure to air pollutants have had inconsistent results. Ghosh et al in California revealed that prenatal exposure to traffic-related air pollution was associated with the risk of ALL [32]. A population-based study from Canada suggested that exposure to ambient air pollution during the first trimester of pregnancy, may cause astrocytoma and ALL [33]. However, Peckham-Gregory et al. conducted a population-based case-control study in Texas to evaluate the association between maternal

residential proximity to major roadways and developing ALL and AML in children, and reached different results. They indicated that mothers who lived closer than 500 meters to a major roadway and mothers who lived in high roadway density areas were not more likely to have a child with ALL or AML [34]. These controversial results indicate that there is a need for more high-quality studies with higher sample sizes, and better control of confounding to determine the role of air pollution exposure in the development of childhood leukemia.

Strengths and limitation

We estimated the simultaneous effect of several different air pollutants on the incidence of leukemia and we also presented a holistic picture of the effect of complex mixtures of air pollution on childhood leukemia. However, our study indeed has limitations. We did not use individual-level data; thus, we were not able to control for confounding variables at individual-level, such as blood group, family history of cancer, taking medicine during pregnancy, father's job, history of radiation, smoking, and other factors such as genetics, nutrition status, cultural context, and behavioral patterns. However, associations between these unmeasured individual factors with ambient air pollutants would need to be rather strong and consistent across the city to effectively bias our estimates. Moreover, we included potential confounding covariates at area level based on administrative survey data. The other major limitation relates to our estimates of exposure. First, our study used recent air pollution data rather than the past life-time estimates thus, our estimates may represent the true but unknown associations only if current and past conditions were highly spatially correlated and if time trends were similar in all areas. We have no data to evaluate this. If trends over time differed in dependence of the concentrations in the past, differential biases of the estimates could result. For example, if areas with particularly high pollution in the past had seen stronger decreases than areas with lower pollution, associations between recent levels of exposure and current incidence of leukemia could be inflated. Second, we assigned the same concentrations to all people living in the same area. Although this area level mean values may well reflect the average home outdoor concentration of the area specific inhabitants, our approach ignored within-area variation, which has been demonstrated in previous studies in Tehran [10-12]. Our method may most likely – at least in theory – be subject to Berkson error [35-37], thus, point estimates may not be biased.

Conclusion

This is the first study to examine the association between multiple air pollutants and leukemia incidence in Iran. Our findings suggest that exposure to VOCs and nitrogen oxides may be associated with increased leukemia incidence in Tehran. Further research with individual data and better control of confounding covariates is needed to confirm the role of air pollution in childhood leukemia development.

List of abbreviations

LPA: Latent profile analysis; IRR: Incidence Rate Ratios; HDI: Human Development Index; ASR: Age Standardized Rates; LUR: Land Use Regression; VOCs: Volatile Organic Compounds; AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion; LMR-LRT: Lo-Mendell-Rubin Likelihood Ratio Test; CI: Confidence Intervals; NB: Negative Binomial; ALL: Acute Lymphoblastic Leukemia; AML: Acute Myeloid Leukemia.

Declarations

Acknowledgements

The authors thank the Cancer Department of the Iranian Ministry of Health, and the Tehran Air Quality Control Company (AQCC), which provided that data for this project. HA is supported by Novo Nordisk Foundation Challenge Program: Harnessing the Power of Big Data to Address the Societal Challenge of Aging (NNF17OC0027812).

Authors' contributions

Z.K., N.K. and M.R study design and Z.K., N.K., M.P conceptualization and M.R, Z.K. and M.P data cleaning and K.E., M.E.A, H.A., S.M.T.-S., N.K data management and statistical analysis and statistical inference and Z.K., M.R., M.P. and H.A Z.K., M.E, N.K. and M.R writing the original draft and Z.K., N.K., M.P., M.R., M.E.A, K.E., S.M.T.-S., N.K., H.A review and editing. All authors participated in revision of the final draft and agreed on the final content.

Funding

This study was financially supported by the Cancer Research Center of Shahid Beheshti University of Medical Sciences, Tehran by Grant No. 25544. HA is supported by Novo Nordisk Foundation Challenge Programme: Harnessing the Power of Big Data to Address the Societal Challenge of Aging [NNF17OC0027812].

Competing interests

The authors declare no competing interests.

Patient consent for publication

Not required

Availability of data and materials

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.

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. As the data was inquired in aggregated form and anonymously, informed consent from individuals or their family was not required. This project was approved by the Ethics in Research Committee of Shahid Beheshti University of Medical Sciences. Ethics Code: IR.SBMU.CRC.REC.1400.003.

Consent for publication

Not applicable.

Fig. 1. Spatial distribution of Leukemia cancer patients (number of cases in 100,000) in different areas of Tehran in 2014-2016 (The labels on the map are district numbers).

Fig. 2. Spearman correlation matrix of air pollutants in the 22 districts of Tehran, Iran, in 2014-2016.

Fig. 3. Standard mean values of pollutants in the two latent profiles in different areas of Tehran, Iran, in 2014-2016.

References:

1. Lien, S.A., et al., *Meta-prediction of MTHFR gene polymorphism-mutations, air pollution, and risks of leukemia among world populations.* Oncotarget, 2017. 8(3): p. 4387-4398.
2. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021.
3. Roshandel, G., et al., *Cancer incidence in Iran in 2014: results of the Iranian National Population-based Cancer Registry.* Cancer epidemiology, 2019. 61: p. 50-58.
4. Turner, M.C., et al., *Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations.* CA Cancer J Clin, 2020.
5. Ou, J.Y., A.C. Kirchhoff, and H.A. Hanson, *Air Pollution across the Cancer Continuum: Extending Our Understanding of the Relationship between Environmental Exposures and Cancer.* Cancer Epidemiol Biomarkers Prev, 2020. 29(10): p. 1876-1879.
6. Filippini, T., et al., *Association between Outdoor Air Pollution and Childhood Leukemia: A Systematic Review and Dose-Response Meta-Analysis.* Environ Health Perspect, 2019. 127(4): p. 46002.
7. Khorrami, Z., et al., *Multiple air pollutant exposure and lung cancer in Tehran, Iran.* Scientific Reports, 2021. 11(1): p. 1-11.
8. Ghaedrahmati, S. and M. Alian, *Health risk assessment of relationship between air pollutants' density and population density in Tehran, Iran.* Human and Ecological Risk Assessment: An International Journal, 2019. 25(7): p. 1853-1869.
9. *World Population Review, Tehran Population 2021. https://worldpopulationreview.com/world-cities/tehran-population. Accessed 22 Apr 2021.*
10. 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.* Science of the total environment, 2014. 488: p. 343-353.
11. Amini, H., et al., *Annual and seasonal spatial models for nitrogen oxides in Tehran, Iran.* Scientific reports, 2016. 6(1): p. 1-11.
12. Amini, H., et al., *Land use regression models for Alkylbenzenes in a middle eastern megacity: Tehran study of exposure prediction for environmental Health Research (Tehran SEPEHR).* Environmental science & technology, 2017. 51(15): p. 8481-8490.

13. 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.* Medical journal of the Islamic Republic of Iran, 2013. 27(4): p. 236.
14. Sadeghi, R. and N. Zanjari, *The inequality of development in the 22 districts of Tehran metropolis.* Social Welfare Quarterly, 2017. 17(66): p. 149-184.
15. Mindrila, D.L., *A typology of child school behavior: Investigation using latent profile analysis and cluster analysis.* Psychology in the Schools, 2016. 53(5): p. 471-487.
16. Nylund, K.L., T. Asparouhov, and B.O. Muthén, *Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study.* Structural equation modeling: A multidisciplinary Journal, 2007. 14(4): p. 535-569.
17. Pourvakhshoori, N., et al., *The association between air pollution and cancers: controversial evidence of a systematic review.* Environ Sci Pollut Res Int, 2020. 27(31): p. 38491-38500.
18. Gao, Y., et al., *Quantitative assessments of indoor air pollution and the risk of childhood acute leukemia in Shanghai.* Environ Pollut, 2014. 187: p. 81-9.
19. Cong, X., *Air pollution from industrial waste gas emissions is associated with cancer incidences in Shanghai, China.* Environ Sci Pollut Res Int, 2018. 25(13): p. 13067-13078.
20. Smith, M.T., *Advances in understanding benzene health effects and susceptibility.* Annual review of public health, 2010. 31: p. 133-148.
21. Whysner, J., et al., *Genotoxicity of benzene and its metabolites.* Mutation Research/Reviews in Mutation Research, 2004. 566(2): p. 99-130.
22. Mondrala, S. and D.A. Eastmond, *Topoisomerase II inhibition by the bioactivated benzene metabolite hydroquinone involves multiple mechanisms.* Chemico-biological interactions, 2010. 184(1-2): p. 259-268.
23. Yang, J., et al., *PTEN methylation involved in benzene-induced hematotoxicity.* Experimental and molecular pathology, 2014. 96(3): p. 300-306.
24. Koehler, C., et al., *Nitrogen dioxide is genotoxic in urban concentrations.* Inhalation toxicology, 2013. 25(6): p. 341-347.
25. Kampa, M. and E. Castanas, *Human health effects of air pollution.* Environmental pollution, 2008. 151(2): p. 362-367.
26. Fiebelkorn, S. and C. Meredith, *Estimation of the leukemia risk in human populations exposed to benzene from tobacco smoke using epidemiological data.* Risk Analysis, 2018. 38(7): p. 1490-1501.
27. Janitz, A.E., et al., *Benzene and childhood acute leukemia in Oklahoma.* Environ Res, 2017. 158: p. 167-173.
28. Ribeiro, A.G., et al., *Residential traffic exposure and lymphohematopoietic malignancies among children in the city of São Paulo, Brazil: An ecological study.* Cancer Epidemiol, 2021. 70: p. 101859.
29. Raaschou-Nielsen, O., et al., *Traffic-related air pollution and risk for leukaemia of an adult population.* Int J Cancer, 2016. 138(5): p. 1111-7.
30. Winters, N., et al., *Exposure to ambient air pollution in Canada and the risk of adult leukemia.* Science of the Total Environment, 2015. 526: p. 153-176.
31. Taj, T., et al., *Exposure to PM(2.5) constituents and risk of adult leukemia in Denmark: A population-based case-control study.* Environ Res, 2020: p. 110418.
32. Ghosh, J.K., et al., *Prenatal exposure to traffic-related air pollution and risk of early childhood cancers.* Am J Epidemiol, 2013. 178(8): p. 1233-9.
33. Lavigne, É., et al., *Maternal exposure to ambient air pollution and risk of early childhood cancers: A population-based study in Ontario, Canada.* Environ Int, 2017. 100: p. 139-147.
34. Peckham-Gregory, E.C., et al., *Maternal Residential Proximity to Major Roadways and the Risk of Childhood Acute Leukemia: A Population-Based Case-Control Study in Texas, 1995-2011.* Int J Environ Res Public Health, 2019. 16(11).

35. Berkson, J., *Are there Two Regressions?* Journal of the American Statistical Association, 1950. 45(250): p. 164-180.
36. Künzli, N. and I.B. Tager, *The semi-individual study in air pollution epidemiology: a valid design as compared to ecologic studies.* Environmental Health Perspectives, 1997. 105(10): p. 1078-1083.
37. Sheppard, L., et al., *Confounding and exposure measurement error in air pollution epidemiology.* Air Quality, Atmosphere & Health, 2012. 5(2): p. 203-216.

Table 1. Description of characteristics of patients, air pollution, and district-level covariates.

Individual-Level variables	
Number of Leukemia patients	3,237
Age of patients (year)	
Median (1st-3st quartile range)	55 (16-69.1)
Gender of patients (male)	
Percent (number)	59.7 (1934)
Grading	
I	4.8(156)
II	33.2(1076)
III	2.8(90)
IV	1(32)
Unknown	58.2(1883)
Topography	
Lymphoid	68.9(2231)
Myeloid	31.1(1006)
air pollutants	Median (1^{st}-3^{st} quartile range)
PM_{10} ($\mu g/m^3$)	101.32(82.35-123.82)
SO_2 (ppb)	52.42(26.38-77.19)
NO_2 (ppb)	46.43(39.95-55.74)
NO (ppb)	78.13(44.47-118.84)
NO_X (ppb)	112.12(83.24-158.49)
benzene ($\mu g/m^3$)	8.12(7.02-9.85)
toluene ($\mu g/m^3$)	24.96(20.85-29.45)
ethylbenzene ($\mu g/m^3$)	5.90(4.97-6.94)
p-xylene ($\mu g/m^3$)	5.71(4.88-6.52)
o-xylene ($\mu g/m^3$)	5.84(4.86-7.46)
m-xylene ($\mu g/m^3$)	10.71(9.06-12.81)
TBTEX ($\mu g/m^3$)	60.57(52.29-70.15)
District level variables	Median (1^{st}-3^{st} quartile range)
Number of districts	22
Total population (persons)	319,443 (263,536-421,225)
Urban green space, per capita (m^2 per 1000 people)	13.9 (10.7-18)
Life expectancy (year)	75.9 (75.4-76.7)
Socioeconomic status *	49.95 (44.21-53.34)

* 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.

Table 2. Fit indices for different latent profile models with number of profiles ranging from 2 to 5.

Profile	Log-likelihood	AIC	BIC	aBIC	LMR LRT	VLMR LRT	entropy
2 profile	**-35505.2**	**71084.5**	**71309.6**	**71192**	**13725.8***	**13856.4***	**0.908**
3 profile	-31630.4	63360.8	63665	63506.1	7676.6	7749.6	0.941
4 profile	-29593.4	59312.9	59696.1	59495.9	4035.5	4073.9	0.951
5 profile	-27590.1	55332.2	55794.5	55553	3968.9	4006.7	0.968

Note: AIC Akaike's Information Criterion, BIC Bayesian Information Criterion, aBIC Sample Size Adjusted BIC, and LMR, Lo–Mendell–Rubin likelihood ratio test; VLMR, Vuong-Lo-Mendell-Rubin likelihood ratio test *P value <0.001.

Models with a significant Lo–Mendell–Rubin likelihood ratio test and Vuong-Lo-Mendell-Rubin likelihood ratio test were preferred. The bolded fit indices indicate that profile 2 was the optimal latent profile model.

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

pollutant	profile	Mean	SD	T	P-value
PM10 (µg/m3)	profile1	87.2	38.3	-21.6	<0.001
	profile2	114.2	31.1		
SO2 (ppb)	profile1	54.8	45.7	-0.32	0.74
	profile2	55.3	53.7		
NO2 (ppb)	profile1	48.4	18.2	-3.84	<0.001
	profile2	50.7	15.7		
NO (ppb)	profile1	68.9	51.6	-19.8	<0.001
	profile2	108.9	63.5		
NOX (ppb)	profile1	91.6	38.8	-33.55	<0.001
	profile2	163.8	80.7		
Benzene (µg/m3)	profile1	8.9	3.2	-26.3	<0.001
	profile2	10.1	3.6		
Toluene (µg/m3)	profile1	19.6	3.5	-56.9	<0.001
	profile2	30.2	6.8		
Ethylbenzene (µg/m3)	profile1	4.7	1.1	-17.4	<0.001
	profile2	7.3	6.1		
P-xylene (µg/m3)	profile1	4.7	0.6	-49.9	<0.001
	profile2	6.7	1.5		
O-xylene (µg/m3)	profile1	4.8	0.8	-53.8	<0.001
	profile2	7.4	1.7		
M-xylene (µg/m3)	profile1	8.6	1.5	-57.1	<0.001
	profile2	13.2	2.9		
TBTEX (µg/m3)	profile1	49.5	7.8	-53.1	<0.001
	profile2	72.9	16.5		

Table 4. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on Leukemia incidence across the districts of Tehran.

	Model 1		Model 2		Model 3	
	IRR (95% CI)	P-value	IRR (95% CI)	P-value	IRR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM_{10} (µg/m3)	0.95(0.87-1.05)	0.352	0.92(0.82-1.03)	0.165	1.03(0.95-1.12)	0.376
Annual SO_2 (ppb)	0.99(0.89-1.09)	0.901	0.99(0.98-1.01)	0.775	0.97(0.91-1.05)	0.575
Annual NO_2 (ppb)	**1.35(1.11-1.64)**	**0.002**	**1.35(1.11-1.64)**	**0.002**	1.19(0.99-1.43)	0.062
Annual NO (ppb)	1.02(0.98-1.07)	0.238	1.03(0.97-1.08)	0.235	1.02(0.99-1.06)	0.130
Annual NO_X (ppb)	**1.04(1.01-1.08)**	**0.003**	**1.07(1.03-1.11)**	**<0.001**	**1.03(1.003-1.06)**	**0.029**
Benzene (µg/m3)	0.97(0.85-1.11)	0.696	0.97(0.85-1.11)	0.965	0.92(0.83-1.02)	0.119
Toluene (µg/m3)	1.11(0.78-1.56)	0.550	1.14(0.71-1.85)	0.570	1.11(0.85-1.44)	0.436
Ethylbenzene (µg/m3)	1.08(0.27-4.28)	0.906	0.97(0.19-4.91)	0.974	1.09(0.38-3.16)	0.865
P-xylene (µg/m3)	1.23(0.14-10.23)	0.844	1.002(0.04-22.77)	0.999	1.24(0.23-6.42)	0.797
O-xylene (µg/m3)	0.48(0.09-2.57)	0.397	0.29(0.04-2.03)	0.214	0.77(0.21-2.84)	0.705
M-xylene (µg/m3)	1.14(0.48-2.72)	0.760	1.11(0.28-4.36)	0.876	1.20(0.61-2.34)	0.589
TBTEX (µg/m3)	1.04(0.89-1.23)	0.561	1.07(0.84-1.36)	0.575	1.03(0.91-1.17)	0.564
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.003(0.99-1.009)	0.227	1.005(0.99-1.01)	0.168	1.003(0.99-1.007)	0.168

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m^2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, life expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 5. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on Myeloid Leukemia incidence in the districts of Tehran.

	Model 1		Model 2		Model 3	
	RR (95% CI)	P-value	RR (95% CI)	P-value	RR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM10 (µg/m3)	0.97(0.88-1.06)	0.547	0.94(0.82-1.08)	0.542	1.01(0.89-1.14)	0.846
Annual SO2 (ppb)	1.004(0.88-1.13)	0.949	1.008(0.88-1.15)	0.904	1.01(0.91-1.11)	0.814
Annual NO2 (ppb)	**1.35(1.11-1.63)**	**0.002**	**1.35(1.11-1.63)**	**0.002**	1.22(0.98-1.52)	0.068
Annual NO (ppb)	1.02(0.98-1.07)	0.260	1.03(0.98-1.09)	0.172	0.99(0.94-1.05)	0.966
Annual NOX (ppb)	0.96(0.93-1.001)	0.057	**0.95(0.92-0.99)**	**0.035**	0.97(0.939-1.01)	0.240
Benzene (µg/m3)	1.71(0.44-6.66)	0.436	4.93(0.60-40.17)	0.136	1.46(0.38-5.47)	0.574
Toluene (µg/m3)	1.19(0.83-1.71)	0.334	1.71(0.97-3.02)	0.061	1.23(0.85-1.77)	0.254
Ethylbenzene (µg/m3)	1.52(0.27-8.53)	0.634	4.59(0.25-82.81)	0.301	1.68(0.31-9.03)	0.545
P-xylene (µg/m3)	1.08(0.87-1.35)	0.454	23.75(0.55-102.3)	0.099	1.78(0.18-17.66)	0.620
O-xylene (µg/m3)	0.81(0.15-4.31)	0.804	0.77(0.06-8.66)	0.837	0.71(0.13-3.87)	0.701
M-xylene (µg/m3)	1.63(0.61-4.31)	0.324	3.03(0.78-11.69)	0.107	1.83(0.74-4.52)	0.190
TBTEX (µg/m3)	1.11(0.92-1.31)	0.255	**1.41(1.07-1.85)**	**0.014**	1.09(0.91-1.32)	0.314
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.003(0.99-1.01)	0.265	1.003(0.99-1.01)	0.259	1.004(0.99-1.01)	0.113

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, Life Expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 6. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on Lymphoid Leukemia incidence in the districts of Tehran.

	Model 1		Model 2		Model 3	
	RR (95% CI)	P-value	RR (95% CI)	P-value	RR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM10 (µg/m3)	0.94(0.87-1.02)	0.207	0.91(0.83-1.01)	0.096	0.99(0.91-1.07)	0.822
Annual SO2 (ppb)	1.08(0.96-1.21)	0.181	1.08(0.96-1.22)	0.172	1.03(0.95-1.12)	0.428
Annual NO2 (ppb)	**1.29(1.07-1.55)**	**0.006**	**1.29(1.07-1.54)**	**0.006**	1.07(0.90-1.27)	0.425
Annual NO (ppb)	1.03(0.98-1.08)	0.157	1.04(0.99-1.10)	0.101	**1.04(0.99-1.08)**	**0.052**
Annual NOX (ppb)	**1.04(1.01-1.08)**	**0.010**	**1.06(1.02-1.10)**	**0.002**	1.01(0.98-1.05)	0.310
Benzene (µg/m3)	1.21(0.36-4.01)	0.754	1.33(0.27-6.48)	0.722	0.46(0.17-1.24)	0.129
Toluene (µg/m3)	1.01(0.72-1.41)	0.949	1.006(0.65-1.54)	0.978	0.92(0.71-1.21)	0.585
Ethylbenzene (µg/m3)	0.84(0.33-2.13)	0.723	0.78(0.27-2.24)	0.658	0.69(0.32-1.46)	0.336
P-xylene (µg/m3)	0.41(0.04-3.51)	0.422	0.21(0.01-3.13)	0.258	0.28(0.05-1.42)	0.126
O-xylene (µg/m3)	0.36(0.08-1.57)	0.175	0.23(0.04-1.26)	0.091	0.32(0.10-1.03)	0.058
M-xylene (µg/m3)	0.83(0.31-2.16)	0.705	0.64(0.16-2.55)	0.533	0.58(0.28-1.20)	0.144
TBTEX (µg/m3)	1.001(0.85-1.17)	0.988	0.99(0.80-1.23)	0.971	0.95(0.84-1.08)	0.433
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.002(0.99-1.009)	0.427	1.003(0.99-1.01)	0.343	0.99(0.99-1.004)	0.884

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space, per capita (m2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status Life Expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

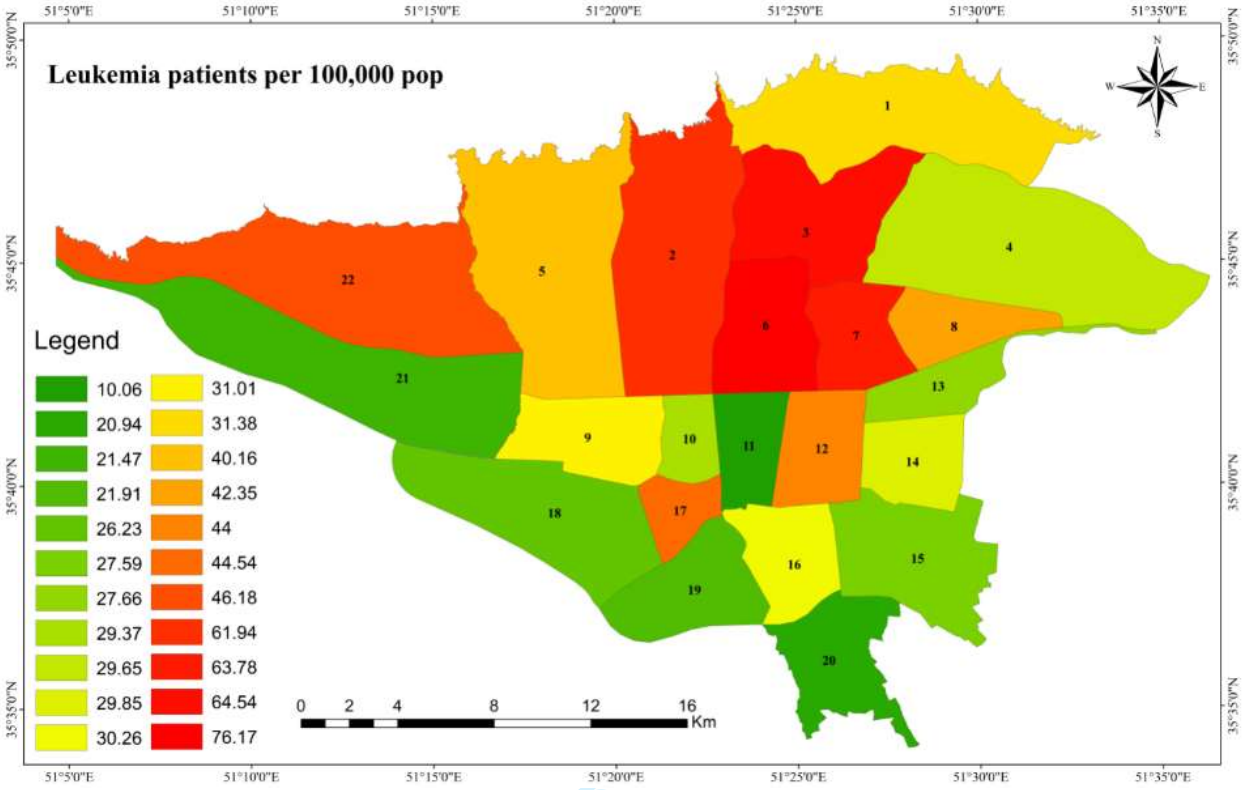

Fig. 1. Spatial distribution of Leukemia cancer patients (number of cases in 100,000) in different areas of Tehran in 2014-2016 (The labels on the map are district numbers).

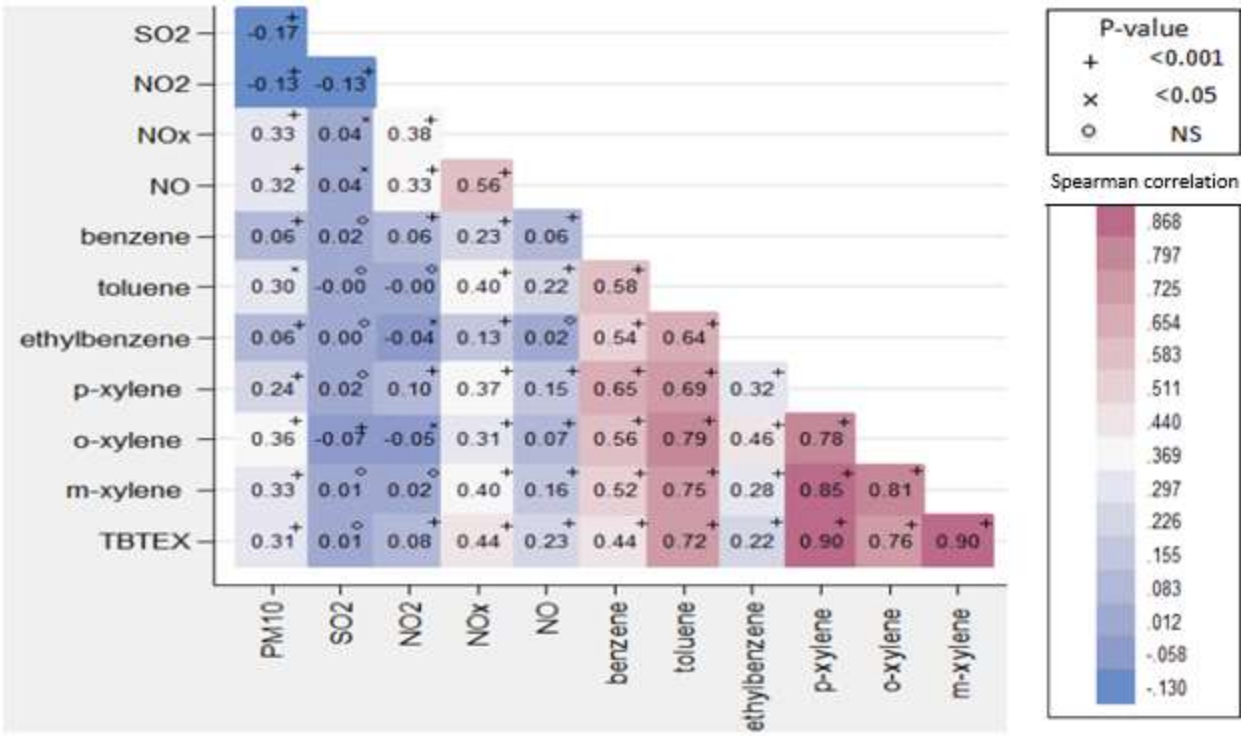

Fig. 2. Spearman correlation matrix of air pollutants in the 22 districts of Tehran, Iran, in 2014-2016.

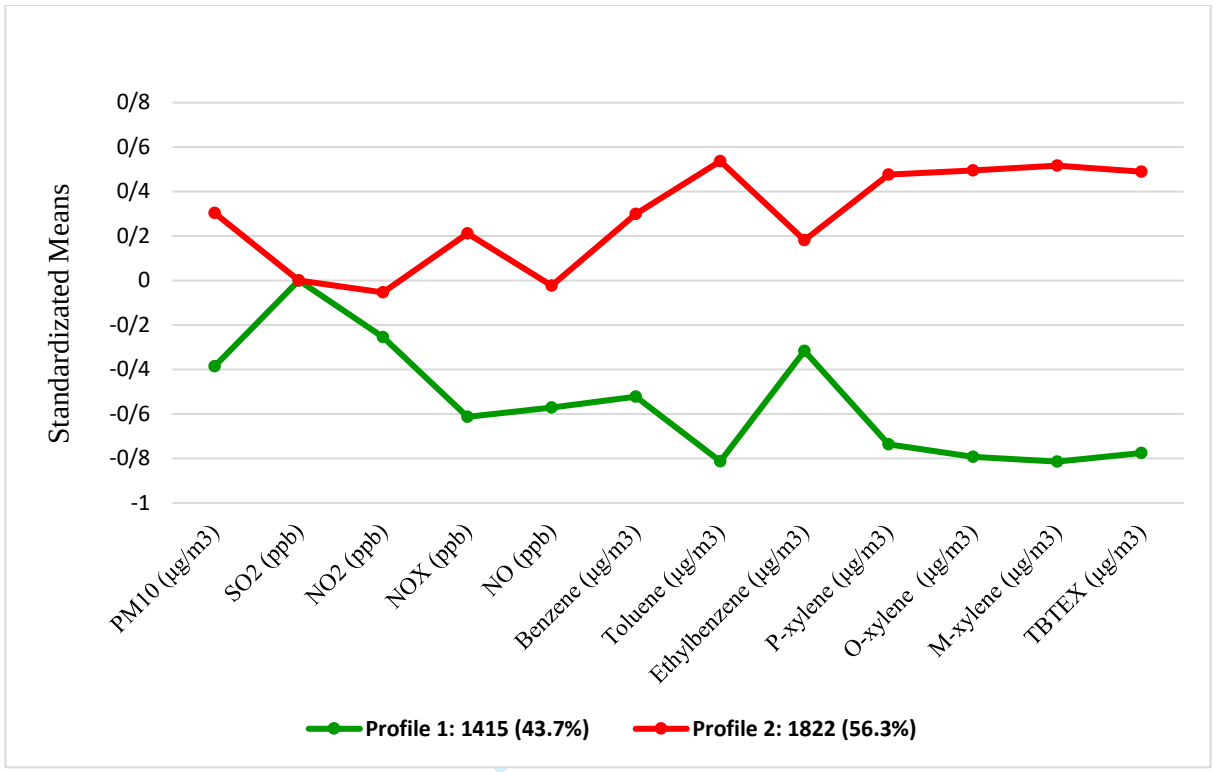

Fig. 3. Standard mean values of pollutants in the two latent profiles in different areas of Tehran, Iran, in 2014-2016.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4,5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4,5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4,5*
Bias	9	Describe any efforts to address potential sources of bias	5
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5,6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5,6
		(*b*) Describe any methods used to examine subgroups and interactions	5,6
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	NA
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	6
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*NA*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	*NA*
		Cross-sectional study—Report numbers of outcome events or summary measures	*6, 16*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	22
		(*b*) Report category boundaries when continuous variables were categorized	16
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	22,23,24
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	23,24
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8,9
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	11

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Multiple Air Pollutants Exposure and Leukemia Incidence in Tehran, Iran

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-060562.R1
Article Type:	Original research
Date Submitted by the Author:	21-Apr-2022
Complete List of Authors:	khorrami, zahra; Kerman University of Medical Sciences, Epidemiology & Biostatistics Pourkhosravani, Mohsen; Shahid Bahonar University of Kerman Eslahi, Marzieh; Kerman University of Medical Sciences, Department of Epidemiology and Biostatistics Rezapour, Maysam; Mazandaran University of Medical Sciences Akbari, Mohammad Esmail; Shahid Beheshti University of Medical Sciences Amini, Heresh; University of Copenhagen Taghavi-Shahri, Seyed Mahmood; University of Copenhagen, Department of Public Health Künzli, Nino ; Schweizerisches Tropen- und Public Health-Institut Etemad, Koorosh; Shahid Beheshti University of Medical Sciences, Department of Epidemiology Khanjani, Narges ; Shahid Beheshti University of Medical Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health
Keywords:	Leukaemia < HAEMATOLOGY, EPIDEMIOLOGY, Epidemiology < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Multiple Air Pollutants Exposure and Leukemia Incidence in Tehran, Iran

Zahra Khorrami[1], Mohsen Pourkhosravani[2], Marzieh Eslahi[3], Maysam Rezapour[4], Mohammad Esmail Akbari[5], Heresh Amini[6], Seyed Mahmood Taghavi-Shahri[7], Nino Künzli[8, 9], Koorosh Etemad[10*], Narges Khanjani[11, 12*]

1. Ophthalmic Research Center, Research Institute for Ophthalmology and Vision Science, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Department of Geography and Urban Planning, Shahid Bahonar University of Kerman, Kerman, Iran
3. Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran
4. Amol Faculty of Paramedical Sciences, Mazandaran University of Medical Sciences, Sari, Iran.
5. Cancer Research Center (CRC), Shahid Beheshti University of Medical Sciences, Tehran, Iran.
6. Department of Public Health, University of Copenhagen, Copenhagen, Denmark
7. Section of Environmental Health, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
8. Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland
9. University of Basel, Basel, Switzerland
10. Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran
11. Environmental Health Engineering Research Center, Kerman University of Medical Sciences, Kerman, Iran
12. Monash Centre for Occupational & Environmental Health, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

*Corresponding authors:
1. Prof. Narges Khanjani, Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran.
Tel/Fax: 034-3132-5102 Email: n_khanjani@kmu.ac.ir.
2. Koorosh Etemad, Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
Tel/Fax: 021- 2243-2040 Email: etemadk@gmail.com

Abstract

Objective: Leukemia is one of the most common cancers and may be associated with exposure to environmental carcinogens, especially outdoor air pollutants. The objective of this study was to investigate the association of ambient air pollution and leukemia in Tehran, Iran.

Design: In this retrospective cohort study, data about the residential district of leukemia cases diagnosed from 2010 to 2016 were inquired from the Ministry of Health cancer database. Data from a previous study was used to determine long-term average exposure to different air pollutants in 22 districts of Tehran. Latent profile analysis (LPA) was used to classify pollutants in two exposure profiles. The association between air pollutants and leukemia incidence was analyzed by negative binomial regression.

Setting: Twenty-two districts of Tehran megacity.

Participants: Leukemia patients.

Outcome measures: The outcome variables were incidence rate ratios of Acute Myeloid and Lymphoid Leukemia across the districts of Tehran.

Results:

The districts with higher concentrations for all pollutants were near the city center. The incidence rate ratio (IRR) was positive but non-significant for most of the air pollutants. However, annual mean NOx was directly and significantly associated with total leukemia incidence in the fully adjusted model (IRR (95% CI): 1.03 (1.003, 1.06) per 10 ppb increase). Based on LPA, districts with a higher multiple air-pollutants profile were also associated with higher leukemia incidence (IRR (95% CI): 1.003 (0.99, 1.007) per 1 ppb increase).

Conclusions:

Our study shows that districts with higher air pollution (nitrogen oxides and multi-pollutants) have higher incidence rates of leukemia in Tehran, Iran. This study warrants conducting further research with individual human data and better control of confounding.

Keywords

Air pollution ; multiple pollutants; pollutant profile; Leukemia; Latent profile analysis

Strengths and limitations of this study

- This is the first study for estimating the simultaneous effect of single and multiple ambient air pollutants on the incidence of leukemia in Iran.
- We included potential confounding covariates at area level based on administrative survey data.
- However, control for confounding variables were not done at the individual level.
- We were not able to adjust for human relocation or migration.

Background

Leukemia is one of the most common hematological malignancies among adults and children [1]. In 2020, leukemia was amongst the 15 more common cancers worldwide and contributed to 2.5 % of all new cases of cancer, and 3.1% of all cancer death. Worldwide, the age standardized incidence and mortality rates of leukemia are 6.3 per 100,000 and 4.0 per 100,000 for men and 4.5 per 100,000 and 2.7 per 100,000 in women, respectively. In countries with low and medium human development index (HDI), the age- standardized incidence and mortality rates of leukemia are, respectively, 3.8 per 100,000 and 3.0 per 100,000 for men, and 2.9 per 100,000 and 2.0 per 100,000 for women [2]. In Iran where cancer is the third leading cause of death after cardiovascular disease, mortality and traffic accidents, leukemia is the 6th most common cancer in men with age standardized rates of 8.32 per 100,000, and the 5th in women with age standardized rates of 5.40 per 100,000 [3].

Outdoor air pollution may have serious adverse effects on human health, including cancer [4]. It has been proven that some environmental factors including ambient air pollution are carcinogenic to humans [5].

The etiology of leukemia remains mostly unknown. Some studies have reported that adverse gene-environment interactions may lead to leukemia [1]. A recent meta-analysis on the association between outdoor air pollution and childhood leukemia showed a positive linear association between benzene exposure and risk of childhood leukemia, particularly for acute myeloid leukemia, among children under 6 years of age, and nitrogen dioxide at high levels was associated with leukemia [6].

In the current study, we aimed to examine the association between estimates of district level averages of multiple ambient air pollutants and acute leukemia incidence across Tehran districts using latent profile analysis (LPA) method. LPA is a robust technique, mainly used to identify subtypes of homogeneous latent classes or subgroups within a large heterogeneous group. This iterative process, clusters similar profiles together to generate distinct subgroups/classes [7].

Methodology

Research location

The present retrospective cohort study was carried out based on annual mean air pollution levels in 22 districts of Tehran megacity, which is the capital of Iran. According to the World Population Review report, Tehran's 2021 population was estimated to be 9.2 million inhabitants. Tehran includes 22 districts in which each include from 174,239 to 919,001 residents according to the latest 2016 census [8]. Population density is higher in the central, western, and southern regions [9]. Tehran suffers from severe ambient air pollution as documented by numerous studies [9-12]. The central districts with higher population densities (five districts of 2, 6, 10, 11, and 12 in central Tehran) have more air pollution [9].

Data Sources

Leukemia data

Information about leukemia patients (acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) residing in Tehran, diagnosed between 2010 and 2016 and their residential address (on a district basis) were obtained from the Ministry of Health's Cancer Registry.

Exposure assessment

The annual mean concentrations of several pollutants in the 22 districts of Tehran were obtained from land use regression (LUR) models developed in previous studies, for PM_{10}, SO_2, NO, NO_2, and NO_X in Tehran, which were based on measurements conducted at 23 regulatory network monitoring sites in 2010 [10, 11]. Data about volatile organic compounds (VOCs) were taken from a previous study [12]. The average of sampling site estimates for each pollutant, in each district was determined and included in the analyzes of this study.

Covariates

District level data including urban green space per capita, life expectancy, and socioeconomic status, was extracted from the Urban Health Equity Assessment and Response Tool (Urban HEART-2), which has been conducted in the 22 districts of Tehran. A detailed description of the Urban HEART-2 study can be found elsewhere [13]**.** The socio-economic indicators of the 22 districts of Tehran was extracted from a study conducted by Sadeghi et al [14].

Statistical analyses

This study used latent profile analysis (LPA) to investigate the patterns of multiple air pollutants exposure. LPA is a statistical method to identify unobserved subgroups (profile) within

populations based on observed variables. LPA has several advantages over traditional methods, such as cluster analysis, e.g. it does not require researchers to specify the number of profiles in advance, which is more suitable for addressing research questions that are exploratory in nature. Moreover, unlike cluster analysis that assigns individuals to clusters absolutely, LPA assigns individuals to subgroups probabilistically [15].

Latent profile analysis (LPA) as a person-centered approach can be used to examine the patterns of multiple air pollutants. LPA is a statistical method for identifying unobserved subgroups within populations based on observed indicators. In contrast to traditional methods, such as cluster analysis, LPA has several advantages. LPA does not require researchers to determine the number of profiles beforehand, and this is more suitable to answer research questions that are exploratory in nature. Also, empirical indicators are available to determine the optimal number of profiles. In addition, LPA allocates individuals to subgroups probabilistically, taking into account the rate of classification uncertainty, and uses multiple statistical indices for determining the optimal number of subgroups.

Most air pollutants in this study had a skewed distribution and were transformed by natural logarithm before LPA, except PM_{10} and SO_2. Several latent profile models were performed, ranging from two to five latent profiles. The most appropriate number of subgroups was identified based on statistical criteria and profile interpretability. The statistical criteria included the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample-size adjusted Bayesian information criterion (a-BIC), the Lo-Mendell-Rubin likelihood ratio test (LMR-LRT), and the Vuong-Lo-Mendell-Rubin likelihood ratio tests (VLMR-LRT). Smaller values of the AIC, BIC, and aBIC indicate a better fit model [16].

A significant p-value of the LMR LRT and VLMR LRT (i.e. $P < 0.05$) indicates a significant improvement in model fit in the k-class model compared to the (k - 1)-class model and thus rejects the (k - 1)-class model and suggests choosing a model with k classes.

In other words, each number of classes (or profiles, shown as k) is compared to the number of classes which is 1 unit less (k-1), and if the VLRT and LMRT values are not significant for higher classes, the model with less classes (k-1) is preferred and will get chosen [17].

Data preparation was done in Stata version 14. LPA was conducted using Mplus version 7.4. One-way ANOVA and post hoc follow-up tests were used to investigate the differences between

profiles of multiple pollutions in terms of each component of air pollution. Management of missing data and other statistical preparation details have been mentioned in our previous publication [8].

Kolmogorov Smirnov test was used to test the normality of the pollutants data and because the data were not normally distributed, Spearman's correlation test was used to estimate the correlation between pollutants. As the number of leukemia 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, and district level (by using them as a covariate). Statistical analyses were performed using Mplus version 7.4, and Stata version 14 (Stata Corp LLC; College Station, TX, USA). ArcGIS version 10.8 was used for exposure assignment to districts and data visualization.

Patient and public involvement

Patients and the public were not involved in the design and conduct of this research.

Results

Summary statistics of patients, air pollutants, and information about the area under study are shown in Table 1. The total number of leukemia cases in 2010–2016 in all districts of Tehran was 3,237. The distribution of leukemia patients in different districts of Tehran is shown in Figure 1. The highest numbers of leukemia patients per 100,000 people were in districts 6, 3, 7 and 2, respectively (Figure 1).

As shown in Figure 2, there was a strong correlation between pollutants, most notably for VOCs (benzene, toluene, m-xylene, p-xylene, o-xylene and TBTEX) (p-value<0.001). The positive correlation between NO and NOx was weaker (r=0.56).

Fit indices for the different LPA models are displayed in Table 2. Several latent profile models were considered. Although, the AIC, BIC, and aBIC of the 2-profile were more than other models, the LMR of LRT (Lo–Mendell–Rubin likelihood ratio test) and VLMR (Vuong-Lo-Mendell-Rubin likelihood ratio test) were significant in the two-profile model.

Figure 3 shows the multi-pollution profiles. Profile 1 had the lowest scores for all pollutants including (PM_{10}, SO2, NO, NO_2, NO_X, benzene, toluene, ethylbenzene, m-xylene, p-xylene, o-xylene and TBTEX). We labeled this profile as "low 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 two profiles, except for SO_2.

Table 4 shows the IRR estimates and 95% confidence interval (CI) by single-pollutant and multiple-pollutant multivariable negative binomial regression models, adjusted for age, gender, socioeconomic status and life expectancy.

In single-pollutant models, NO_2 and NO_X were significantly associated with increased leukemia incidence in model 1 (adjusted for age and gender), and also in model 2 (adjusted in addition for urban green space per capita) with IRR (95% CI) of model 2, respectively, 1.35 (1.11-1.64) and 1.07 (1.03-1.11) per 10 ppb increase in pollutants. In model 3, after adjustment for age, gender, socioeconomic status, and life expectancy, only NO_X was significantly associated with increased leukemia incidence with an IRR (95% CI) of 1.03 (1.003-1.06) per 10 ppb increase in NOx. However, NO_2 was borderline significantly associated with increased leukemia incidence (IRR=1.19, CI 95%=0.99-1.43) in this model.

In multi-pollutant models, the high multiple-air-pollutants profile was associated with higher leukemia incidence when compared with the low multiple-air-pollutants profile, but this association did not reach significance.

Tables 5 and 6 show the IRR (95% CI) estimates by single-pollutant and multiple-pollutant multivariable negative binomial regression models for acute myeloid and lymphoid leukemia incidence, respectively. NO, NO_2 and NO_x were related to acute increased lymphoid leukemia incidence, while NO_2 and TBTEX were related to increased acute myeloid leukemia incidence.

Tables 7 and 8 show the IRR (95% CI) estimates by single-pollutant and multiple-pollutant multivariable negative binomial regression models for total acute leukemia incidence, respectively in children and adults. Increase in all single pollutants except SO_2 and high multiple pollutants were related to increased acute leukemia incidence in children (≤14 years old), while nitrogen oxides, toluene, ethylbenzene, o-xylene and TBTEX were related to increased acute leukemia incidence among adults.

Discussion

The findings of our study show that exposure to air pollutants, such as nitrogen oxides and VOCs may be associated with the incidence of leukemia. Our study was the first to investigate the effect of single and multiple ambient air pollutants on leukemia in Iran.

Leukemia is one of the most common cancers in children and adults. The cause of leukemia is currently unknown [18]. However, some sources have suggested that genetic and transgenic mutations due to environmental factors may contribute to leukemia [1, 19]. A study in Shanghai, China showed that air pollution from industrial waste gas emissions was associated with the incidence of several cancers including leukemia [20].

Many experimental studies have shown that air pollutants may be carcinogenic by causing DNA damage and mutations [21-26].

A review indicated that exposure to air pollutants is associated with leukemia stronger than other cancers [18]. Carlos-Wallace et al. also conducted a meta-analysis and reported associations between childhood leukemia and benzene exposure. They indicated that in studies that evaluated benzene versus all solvents, the pooled RR (Relative Risk from meta-analysis) was higher for maternal exposure compared to paternal exposure. For NO_2, an excess risk was reported in concentration-response meta-analysis from 40 $\mu g/m^3$ to 60 $\mu g/m^3$; however, the increase was not statistically significant and was mainly related to ALL [6].

Our study showed some associations between benzene and leukemia incidence among children. Similarly, a study in the United Kingdom revealed that there was an increased risk of leukemia from low-level exposure to benzene from smoking, and that benzene may contribute to up to a third of smoking-induced leukemia. The contribution of benzene to smoking-induced leukemia was estimated to be between 9% and 24%. For AML this contribution was estimated as 11-30% [27]. Janitz et al. conducted a case-control study between 1997 and 2012 including 307 patients with acute leukemia from the Oklahoma Central Cancer Registry and 1013 controls matched on week of birth from birth certificates and showed no association between benzene and childhood leukemia, in Oklahoma. However, based on the analysis of quartiles they indicated that there was a positive concentration-response relation between benzene and AML [28].

Our study indicated that NO_2 increased the incidence of acute lymphoid leukemia, and total leukemia among children and adults. Similarly, findings of the population-based study by

Ribeiro et al in São Paulo, Brazil has shown that NO_2 and traffic density were associated with Hodgkin lymphoma and lymphoid leukemia in children; and the variations in the incidence rate ratios in gender and SES groups may be because of differences in underlying risk and exposure profiles [29]. In addition, Raaschou-Nielsen et al. conducted a nationwide case-control study in Denmark and indicated that long-term exposure to traffic-related air pollutants (NO_x and NO_2) was associated with acute myeloid leukemia, but not other subtypes of leukemia, in the general population [30]. Our study revealed that NO, NO_2 and NO_x are related to increased acute lymphoid leukemia incidence, while NO_2 and TBTEX are related to increased acute myeloid leukemia incidence. While, a Canadian population-based case–control study revealed a weak association between all forms of leukemia only at low concentrations of NO_2. The study showed an 'n-shaped' response function between exposure to NO_2 and all forms of leukemia. The OR was 1.20 (95% CI: 0.97-1.48) from the 10^{th} percentile (4.51 ppb) to the median (14.66 ppb), then the OR decreased to 0.79 (95% CI: 0.68, 0.93) from the 75th percentile (22.75) to the 90th (29.7 ppb) range [31]. Some differences in results may be attributable to differences in settings and population characteristics.

In a large nationwide population-based case-control study in Denmark, Taj et al. showed a high risk for overall leukemia in association with secondary inorganic aerosols including nitrate (NO_3), and that AML was associated with NO_3 [32], which was similar to our findings.

Studies focusing on the association between childhood leukemia and exposure to air pollutants have had inconsistent results. Ghosh et al in California revealed that prenatal exposure to traffic-related air pollution was associated with the risk of ALL [33]. A population-based study from Canada suggested that exposure to ambient air pollution during the first trimester of pregnancy, may cause astrocytoma and ALL [34]. However, Peckham-Gregory et al. conducted a population-based case-control study in Texas to evaluate the association between maternal residential proximity to major roadways and developing ALL and AML in children, and reached different results. They indicated that mothers who lived closer than 500 meters to a major roadway and mothers who lived in high roadway density areas were not more likely to have a child with ALL or AML [35].

Although many studies have shown a possible relation between air pollution and leukemia, there is still a need for more high-quality studies with higher sample sizes, and better control of confounders.

Strengths and limitation

In this study, we estimated the simultaneous effect of several different air pollutants on the incidence of leukemia. However, our study had several limitations. We did not use individual-level data; thus, we were not able to control for confounding variables at individual-level, such as blood group, family history of cancer, taking medicine during pregnancy, parents' job, history of radiation, smoking, and other factors such as genetics, nutrition status, cultural context, and behavioral patterns. But we did include potential confounding covariates at regional level.

The other major limitation of this study is our estimates of exposure. First, our study used some more recent air pollution data, not past life-time exposure estimates, because we did not have a better option. In addition, we did not have data on the length of residence of patients in the study regions, and life-time relocation or migration. Second, we assigned the same exposure concentrations to all people living in the same area. Although the mean area level values may well reflect the average exposure levels of the inhabitants, our approach ignored within-area variation, which has been demonstrated in previous studies in Tehran [10-12].

Conclusion

This is the first study to examine the association between multiple air pollutants and leukemia incidence in Iran. Our findings suggest that exposure to VOCs, nitrogen oxides and/or multiple ambient air pollutants may be associated with increased leukemia incidence in Tehran. Further research with individual data and better control of confounding covariates is needed to confirm the role of air pollution in human leukemia.

List of abbreviations

LPA: Latent profile analysis; IRR: Incidence Rate Ratios; HDI: Human Development Index; ASR: Age Standardized Rates; LUR: Land Use Regression; VOCs: Volatile Organic

Compounds; AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion; LMR-LRT: Lo-Mendell-Rubin Likelihood Ratio Test; CI: Confidence Intervals; NB: Negative Binomial; ALL: Acute Lymphoblastic Leukemia; AML: Acute Myeloid Leukemia.

Declarations

Acknowledgements

The authors thank the Cancer Department of the Iranian Ministry of Health, and the Tehran Air Quality Control Company (AQCC), which provided that data for this project. HA is supported by Novo Nordisk Foundation Challenge Program: Harnessing the Power of Big Data to Address the Societal Challenge of Aging (NNF17OC0027812).

Authors' contributions

Z.K., N.K. and M.R study design and Z.K., N.K., M.P conceptualization and M.R, Z.K. and M.P data cleaning and K.E., M.E.A, H.A., S.M.T.-S., N.K data management and statistical analysis and statistical inference and Z.K., M.R., M.P. and H.A Z.K., M.E, N.K. and M.R writing the original draft and Z.K., N.K., M.P., M.R., M.E.A, K.E., S.M.T.-S., N.K., H.A review and editing. All authors participated in revision of the final draft and agreed on the final content.

Funding

This study was financially supported by the Cancer Research Center of Shahid Beheshti University of Medical Sciences, Tehran by Grant No. 25544. HA is supported by Novo Nordisk Foundation Challenge Programme: Harnessing the Power of Big Data to Address the Societal Challenge of Aging [NNF17OC0027812].

Competing interests

The authors declare no competing interests.

Patient consent for publication

Not required

Availability of data and materials

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.

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. As the data was inquired in aggregated form and anonymously, informed consent from individuals or their family was not required. This project was approved by the Ethics in Research Committee of Shahid Beheshti University of Medical Sciences. Ethics Code: IR.SBMU.CRC.REC.1400.003.

Consent for publication

Not applicable.

Fig. 1. Spatial distribution of Leukemia patients (number of cases in 100,000) in different areas of Tehran in 2014-2016 (The labels on the map are district numbers).

Fig. 2. Spearman correlation matrix of air pollutants in the 22 districts of Tehran, Iran, in 2014-2016.

Fig. 3. Standard mean values of pollutants in the two latent profiles in different areas of Tehran, Iran, in 2014-2016.

References:

1. Lien, S.A., et al., *Meta-prediction of MTHFR gene polymorphism-mutations, air pollution, and risks of leukemia among world populations.* Oncotarget, 2017. **8**(3): p. 4387-4398.
2. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021.
3. Roshandel, G., et al., *Cancer incidence in Iran in 2014: results of the Iranian National Population-based Cancer Registry.* Cancer epidemiology, 2019. **61**: p. 50-58.
4. Turner, M.C., et al., *Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations.* CA Cancer J Clin, 2020.
5. Ou, J.Y., A.C. Kirchhoff, and H.A. Hanson, *Air Pollution across the Cancer Continuum: Extending Our Understanding of the Relationship between Environmental Exposures and Cancer.* Cancer Epidemiol Biomarkers Prev, 2020. **29**(10): p. 1876-1879.
6. Filippini, T., et al., *Association between Outdoor Air Pollution and Childhood Leukemia: A Systematic Review and Dose-Response Meta-Analysis.* Environ Health Perspect, 2019. **127**(4): p. 46002.
7. Fernández, R.S., et al., *Psychological distress associated with COVID-19 quarantine: Latent profile analysis, outcome prediction and mediation analysis.* J Affect Disord, 2020. **277**: p. 75-84.
8. Khorrami, Z., et al., *Multiple air pollutant exposure and lung cancer in Tehran, Iran.* Scientific Reports, 2021. **11**(1): p. 1-11.
9. Ghaedrahmati, S. and M. Alian, *Health risk assessment of relationship between air pollutants' density and population density in Tehran, Iran.* Human and Ecological Risk Assessment: An International Journal, 2019. **25**(7): p. 1853-1869.

10. 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.* Science of the total environment, 2014. **488**: p. 343-353.
11. Amini, H., et al., *Annual and seasonal spatial models for nitrogen oxides in Tehran, Iran.* Scientific reports, 2016. **6**(1): p. 1-11.
12. Amini, H., et al., *Land use regression models for Alkylbenzenes in a middle eastern megacity: Tehran study of exposure prediction for environmental Health Research (Tehran SEPEHR).* Environmental science & technology, 2017. **51**(15): p. 8481-8490.
13. 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.* Medical journal of the Islamic Republic of Iran, 2013. **27**(4): p. 236.
14. Sadeghi, R. and N. Zanjari, *The inequality of development in the 22 districts of Tehran metropolis.* Social Welfare Quarterly, 2017. **17**(66): p. 149-184.
15. Mindrila, D.L., *A typology of child school behavior: Investigation using latent profile analysis and cluster analysis.* Psychology in the Schools, 2016. **53**(5): p. 471-487.
16. Nylund, K.L., T. Asparouhov, and B.O. Muthén, *Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study.* Structural equation modeling: A multidisciplinary Journal, 2007. **14**(4): p. 535-569.
17. Tein, J.-Y., S. Coxe, and H. Cham, *Statistical power to detect the correct number of classes in latent profile analysis.* Structural equation modeling: a multidisciplinary journal, 2013. **20**(4): p. 640-657.
18. Pourvakhshoori, N., et al., *The association between air pollution and cancers: controversial evidence of a systematic review.* Environ Sci Pollut Res Int, 2020. **27**(31): p. 38491-38500.
19. Gao, Y., et al., *Quantitative assessments of indoor air pollution and the risk of childhood acute leukemia in Shanghai.* Environ Pollut, 2014. **187**: p. 81-9.
20. Cong, X., *Air pollution from industrial waste gas emissions is associated with cancer incidences in Shanghai, China.* Environ Sci Pollut Res Int, 2018. **25**(13): p. 13067-13078.
21. Smith, M.T., *Advances in understanding benzene health effects and susceptibility.* Annual review of public health, 2010. **31**: p. 133-148.
22. Whysner, J., et al., *Genotoxicity of benzene and its metabolites.* Mutation Research/Reviews in Mutation Research, 2004. **566**(2): p. 99-130.
23. Mondrala, S. and D.A. Eastmond, *Topoisomerase II inhibition by the bioactivated benzene metabolite hydroquinone involves multiple mechanisms.* Chemico-biological interactions, 2010. **184**(1-2): p. 259-268.
24. Yang, J., et al., *PTEN methylation involved in benzene-induced hematotoxicity.* Experimental and molecular pathology, 2014. **96**(3): p. 300-306.
25. Koehler, C., et al., *Nitrogen dioxide is genotoxic in urban concentrations.* Inhalation toxicology, 2013. **25**(6): p. 341-347.
26. Kampa, M. and E. Castanas, *Human health effects of air pollution.* Environmental pollution, 2008. **151**(2): p. 362-367.
27. Fiebelkorn, S. and C. Meredith, *Estimation of the leukemia risk in human populations exposed to benzene from tobacco smoke using epidemiological data.* Risk Analysis, 2018. **38**(7): p. 1490-1501.
28. Janitz, A.E., et al., *Benzene and childhood acute leukemia in Oklahoma.* Environ Res, 2017. **158**: p. 167-173.

29. Ribeiro, A.G., et al., *Residential traffic exposure and lymphohematopoietic malignancies among children in the city of São Paulo, Brazil: An ecological study.* Cancer Epidemiol, 2021. **70**: p. 101859.
30. Raaschou-Nielsen, O., et al., *Traffic-related air pollution and risk for leukaemia of an adult population.* Int J Cancer, 2016. **138**(5): p. 1111-7.
31. Winters, N., et al., *Exposure to ambient air pollution in Canada and the risk of adult leukemia.* Science of the Total Environment, 2015. **526**: p. 153-176.
32. Taj, T., et al., *Exposure to PM(2.5) constituents and risk of adult leukemia in Denmark: A population-based case-control study.* Environ Res, 2020: p. 110418.
33. Ghosh, J.K., et al., *Prenatal exposure to traffic-related air pollution and risk of early childhood cancers.* Am J Epidemiol, 2013. **178**(8): p. 1233-9.
34. Lavigne, É., et al., *Maternal exposure to ambient air pollution and risk of early childhood cancers: A population-based study in Ontario, Canada.* Environ Int, 2017. **100**: p. 139-147.
35. Peckham-Gregory, E.C., et al., *Maternal Residential Proximity to Major Roadways and the Risk of Childhood Acute Leukemia: A Population-Based Case-Control Study in Texas, 1995-2011.* Int J Environ Res Public Health, 2019. **16**(11).

Table 1. Description of characteristics of patients, air pollution, and district-level covariates.

Individual-Level variables	
Number of Leukemia patients	3,237
Age of patients (year)	
Median (1st-3st quartile range)	55 (16-69.1)
Gender of patients (male)	
Percent (number)	59.7 (1934)
Grading	
I	4.8(156)
II	33.2(1076)
III	2.8(90)
IV	1(32)
Unknown	58.2(1883)
Topography	
Lymphoid	68.9(2231)
Myeloid	31.1(1006)
Air pollutants	Median (1^{st}-3^{st} quartile range)
PM_{10} ($\mu g/m^3$)	101.32(82.35-123.82)
SO_2 (ppb)	52.42(26.38-77.19)
NO_2 (ppb)	46.43(39.95-55.74)
NO (ppb)	78.13(44.47-118.84)
NO_X (ppb)	112.12(83.24-158.49)
benzene ($\mu g/m^3$)	8.12(7.02-9.85)
toluene ($\mu g/m^3$)	24.96(20.85-29.45)
ethylbenzene ($\mu g/m^3$)	5.90(4.97-6.94)
p-xylene ($\mu g/m^3$)	5.71(4.88-6.52)
o-xylene ($\mu g/m^3$)	5.84(4.86-7.46)
m-xylene ($\mu g/m^3$)	10.71(9.06-12.81)
TBTEX ($\mu g/m^3$)	60.57(52.29-70.15)
District level variables	Median (1^{st}-3^{st} quartile range)
Number of districts	22
Total population (persons)	319,443 (263,536-421,225)
Urban green space, per capita (m^2 per 1000 people)	13.9 (10.7-18)
Life expectancy (year)	75.9 (75.4-76.7)
Socioeconomic status *	49.95 (44.21-53.34)

* 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.

Table 2. Fit indices for different latent profile models with number of profiles ranging from 2 to 5.

Profile	Log-likelihood	AIC	BIC	aBIC	LMR LRT	VLMR LRT	entropy
2 profile	**-35505.2**	**71084.5**	**71309.6**	**71192**	**13725.8***	**13856.4***	**0.908**
3 profile	-31630.4	63360.8	63665	63506.1	7676.6	7749.6	0.941
4 profile	-29593.4	59312.9	59696.1	59495.9	4035.5	4073.9	0.951
5 profile	-27590.1	55332.2	55794.5	55553	3968.9	4006.7	0.968

Note: AIC Akaike's Information Criterion, BIC Bayesian Information Criterion, aBIC Sample Size Adjusted BIC, and LMR, Lo–Mendell–Rubin likelihood ratio test; VLMR, Vuong-Lo-Mendell-Rubin likelihood ratio test *P value <0.001.

Models with a significant Lo–Mendell–Rubin likelihood ratio test and Vuong-Lo-Mendell-Rubin likelihood ratio test were preferred. The bolded fit indices indicate that profile 2 was the optimal latent profile model.

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

pollutant	profile	Mean	SD	T	P-value
PM10 (µg/m3)	profile1	87.2	38.3	-21.6	<0.001
	profile2	114.2	31.1		
SO2 (ppb)	profile1	54.8	45.7	-0.32	0.74
	profile2	55.3	53.7		
NO2 (ppb)	profile1	48.4	18.2	-3.84	<0.001
	profile2	50.7	15.7		
NO (ppb)	profile1	68.9	51.6	-19.8	<0.001
	profile2	108.9	63.5		
NOX (ppb)	profile1	91.6	38.8	-33.55	<0.001
	profile2	163.8	80.7		
Benzene (µg/m3)	profile1	8.9	3.2	-26.3	<0.001
	profile2	10.1	3.6		
Toluene (µg/m3)	profile1	19.6	3.5	-56.9	<0.001
	profile2	30.2	6.8		
Ethylbenzene (µg/m3)	profile1	4.7	1.1	-17.4	<0.001
	profile2	7.3	6.1		
P-xylene (µg/m3)	profile1	4.7	0.6	-49.9	<0.001
	profile2	6.7	1.5		
O-xylene (µg/m3)	profile1	4.8	0.8	-53.8	<0.001
	profile2	7.4	1.7		
M-xylene (µg/m3)	profile1	8.6	1.5	-57.1	<0.001
	profile2	13.2	2.9		
TBTEX (µg/m3)	profile1	49.5	7.8	-53.1	<0.001
	profile2	72.9	16.5		

Table 4. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on total Leukemia incidence across the districts of Tehran.

	Model 1		Model 2		Model 3	
	IRR (95% CI)	P-value	IRR (95% CI)	P-value	IRR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM_{10} (µg/m3)	0.95(0.87-1.05)	0.352	0.92(0.82-1.03)	0.165	1.03(0.95-1.12)	0.376
Annual SO_2 (ppb)	0.99(0.89-1.09)	0.901	0.99(0.98-1.01)	0.775	0.97(0.91-1.05)	0.575
Annual NO_2 (ppb)	**1.35(1.11-1.64)**	**0.002**	**1.35(1.11-1.64)**	**0.002**	1.19(0.99-1.43)	0.062
Annual NO (ppb)	1.02(0.98-1.07)	0.238	1.03(0.97-1.08)	0.235	1.02(0.99-1.06)	0.130
Annual NO_X (ppb)	**1.04(1.01-1.08)**	**0.003**	**1.07(1.03-1.11)**	**<0.001**	**1.03(1.003-1.06)**	**0.029**
Benzene (µg/m3)	0.97(0.85-1.11)	0.696	0.97(0.85-1.11)	0.965	0.92(0.83-1.02)	0.119
Toluene (µg/m3)	1.11(0.78-1.56)	0.550	1.14(0.71-1.85)	0.570	1.11(0.85-1.44)	0.436
Ethylbenzene (µg/m3)	1.08(0.27-4.28)	0.906	0.97(0.19-4.91)	0.974	1.09(0.38-3.16)	0.865
P-xylene (µg/m3)	1.23(0.14-10.23)	0.844	1.002(0.04-22.77)	0.999	1.24(0.23-6.42)	0.797
O-xylene (µg/m3)	0.48(0.09-2.57)	0.397	0.29(0.04-2.03)	0.214	0.77(0.21-2.84)	0.705
M-xylene (µg/m3)	1.14(0.48-2.72)	0.760	1.11(0.28-4.36)	0.876	1.20(0.61-2.34)	0.589
TBTEX (µg/m3)	1.04(0.89-1.23)	0.561	1.07(0.84-1.36)	0.575	1.03(0.91-1.17)	0.564
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.003(0.99-1.009)	0.227	1.005(0.99-1.01)	0.168	1.003(0.99-1.007)	0.168

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m^2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, life expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 5. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on Acute Myeloid Leukemia incidence in the districts of Tehran.

	Model 1		Model 2		Model 3	
	RR (95% CI)	P-value	RR (95% CI)	P-value	RR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM10 (µg/m3)	0.97(0.88-1.06)	0.547	0.94(0.82-1.08)	0.542	1.01(0.89-1.14)	0.846
Annual SO2 (ppb)	1.004(0.88-1.13)	0.949	1.008(0.88-1.15)	0.904	1.01(0.91-1.11)	0.814
Annual NO2 (ppb)	**1.35(1.11-1.63)**	**0.002**	**1.35(1.11-1.63)**	**0.002**	1.22(0.98-1.52)	0.068
Annual NO (ppb)	1.02(0.98-1.07)	0.260	1.03(0.98-1.09)	0.172	0.99(0.94-1.05)	0.966
Annual NOX (ppb)	0.96(0.93-1.001)	0.057	**0.95(0.92-0.99)**	**0.035**	0.97(0.939-1.01)	0.240
Benzene (µg/m3)	1.71(0.44-6.66)	0.436	4.93(0.60-40.17)	0.136	1.46(0.38-5.47)	0.574
Toluene (µg/m3)	1.19(0.83-1.71)	0.334	1.71(0.97-3.02)	0.061	1.23(0.85-1.77)	0.254
Ethylbenzene (µg/m3)	1.52(0.27-8.53)	0.634	4.59(0.25-82.81)	0.301	1.68(0.31-9.03)	0.545
P-xylene (µg/m3)	1.08(0.87-1.35)	0.454	23.75(0.55-102.3)	0.099	1.78(0.18-17.66)	0.620
O-xylene (µg/m3)	0.81(0.15-4.31)	0.804	0.77(0.06-8.66)	0.837	0.71(0.13-3.87)	0.701
M-xylene (µg/m3)	1.63(0.61-4.31)	0.324	3.03(0.78-11.69)	0.107	1.83(0.74-4.52)	0.190
TBTEX (µg/m3)	1.11(0.92-1.31)	0.255	**1.41(1.07-1.85)**	**0.014**	1.09(0.91-1.32)	0.314
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.003(0.99-1.01)	0.265	1.003(0.99-1.01)	0.259	1.004(0.99-1.01)	0.113

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, Life Expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 6. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on Acute Lymphoid Leukemia incidence in the districts of Tehran.

	Model 1		Model 2		Model 3	
	RR (95% CI)	P-value	RR (95% CI)	P-value	RR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM10 (µg/m3)	0.94(0.87-1.02)	0.207	0.91(0.83-1.01)	0.096	0.99(0.91-1.07)	0.822
Annual SO2 (ppb)	1.08(0.96-1.21)	0.181	1.08(0.96-1.22)	0.172	1.03(0.95-1.12)	0.428
Annual NO2 (ppb)	**1.29(1.07-1.55)**	**0.006**	**1.29(1.07-1.54)**	**0.006**	1.07(0.90-1.27)	0.425
Annual NO (ppb)	1.03(0.98-1.08)	0.157	1.04(0.99-1.10)	0.101	**1.04(0.99-1.08)**	**0.052**
Annual NOX (ppb)	**1.04(1.01-1.08)**	**0.010**	**1.06(1.02-1.10)**	**0.002**	1.01(0.98-1.05)	0.310
Benzene (µg/m3)	1.21(0.36-4.01)	0.754	1.33(0.27-6.48)	0.722	0.46(0.17-1.24)	0.129
Toluene (µg/m3)	1.01(0.72-1.41)	0.949	1.006(0.65-1.54)	0.978	0.92(0.71-1.21)	0.585
Ethylbenzene (µg/m3)	0.84(0.33-2.13)	0.723	0.78(0.27-2.24)	0.658	0.69(0.32-1.46)	0.336
P-xylene (µg/m3)	0.41(0.04-3.51)	0.422	0.21(0.01-3.13)	0.258	0.28(0.05-1.42)	0.126
O-xylene (µg/m3)	0.36(0.08-1.57)	0.175	0.23(0.04-1.26)	0.091	0.32(0.10-1.03)	0.058
M-xylene (µg/m3)	0.83(0.31-2.16)	0.705	0.64(0.16-2.55)	0.533	0.58(0.28-1.20)	0.144
TBTEX (µg/m3)	1.001(0.85-1.17)	0.988	0.99(0.80-1.23)	0.971	0.95(0.84-1.08)	0.433
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.002(0.99-1.009)	0.427	1.003(0.99-1.01)	0.343	0.99(0.99-1.004)	0.884

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space, per capita (m2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status Life Expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 7. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on total Leukemia incidence in children (≤14 years old) across the districts of Tehran.

	Model 1		Model 2		Model 3	
	IRR (95% CI)	P-value	IRR (95% CI)	P-value	IRR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM_{10} (µg/m3)	**0.87(0.84-0.92)**	**<0.001**	**0.78(0.74-0.82)**	**<0.001**	**1.11(1.03-1.18)**	**0.003**
Annual SO_2 (ppb)	**0.95(0.91-0.99)**	**0.036**	0.96(0.92-1.008)	0.106	0.99(0.95-1.04)	0.712
Annual NO_2 (ppb)	**1.45(1.32-1.61)**	**<0.001**	**1.45(1.31-1.60)**	**<0.001**	**1.21(1.08-1.35)**	**0.001**
Annual NO (ppb)	**0.93(0.91-0.96)**	**<0.001**	**0.90(0.88-0.93)**	**<0.001**	**1.05(1.02-1.09)**	**0.005**
Annual NO_X (ppb)	**1.04(1.02-1.07)**	**<0.001**	**1.04(1.01-1.07)**	**0.004**	**1.08(1.05-1.10)**	**<0.001**
Benzene (µg/m3)	0.49(0.23-1.05)	0.068	**0.30(0.14-0.67)**	**0.003**	**5.87(2.43-14.16)**	**<0.001**
Toluene (µg/m3)	**0.75(0.63-0.89)**	**0.001**	**0.63(0.52-0.76)**	**<0.001**	**1.77(1.42-2.22)**	**<0.001**
Ethylbenzene (µg/m3)	**0.17(0.07-0.42)**	**<0.001**	**0.06(0.02-0.16)**	**<0.001**	**21.78(6.02-78.85)**	**<0.001**
P-xylene (µg/m3)	**0.14(0.05-0.43)**	**0.001**	**0.02(0.01-0.08)**	**<0.001**	**17.83(4.22-75.29)**	**<0.001**
O-xylene (µg/m3)	**0.11(0.05-0.24)**	**0.001**	**0.04(0.01-0.08)**	**<0.001**	**19.89(5.19-76.24)**	**<0.001**
M-xylene (µg/m3)	**0.37(0.24-0.58)**	**<0.001**	**0.20(0.12-0.33)**	**<0.001**	**4.08(2.13-7.80)**	**<0.001**
TBTEX (µg/m3)	**0.89(0.82-0.97)**	**0.007**	**0.79(0.73-0.87)**	**<0.001**	**1.29(1.16-1.44)**	**<0.001**
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	**1.003(1.001-1.006)**	**0.001**	**1.004(1.001-1.007)**	**0.016**	**1.008(1.005-1.01)**	**<0.001**

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m^2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, life expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 8. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on total Leukemia incidence in adults across the districts of Tehran.

	Model 1		Model 2		Model 3	
	IRR (95% CI)	P-value	IRR (95% CI)	P-value	IRR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM_{10} (µg/m3)	**0.92(0.90-0.95)**	**<0.001**	**0.89(0.87-0.92)**	**<0.001**	1.01(0.97-1.05)	0.479
Annual SO_2 (ppb)	**0.96(0.94-0.99)**	**0.012**	0.97(0.95-1.01)	0.109	0.98(0.96-1.02)	0.390
Annual NO_2 (ppb)	**1.21(1.14-1.27)**	**<0.001**	**1.19(1.13-1.26)**	**<0.001**	**1.10(1.03-1.18)**	**0.003**
Annual NO (ppb)	**0.98(0.97-0.99)**	**0.028**	**0.97(0.95-0.98)**	**<0.001**	**1.03(1.01-1.05)**	**0.005**
Annual NO_X (ppb)	**1.02(1.00-1.03)**	**0.008**	**1.01(0.99-1.02)**	**0.338**	**1.03(1.01-1.04)**	**<0.001**
Benzene (µg/m3)	0.66(0.42-1.05)	0.078	**0.44(0.27-0.72)**	**0.001**	1.44(0.87-2.40)	0.158
Toluene (µg/m3)	**0.87(0.78-0.97)**	**0.017**	**0.79(0.71-0.89)**	**<0.001**	**1.19(1.05-1.37)**	**0.008**
Ethylbenzene (µg/m3)	**0.36(0.19-0.67)**	**0.001**	**0.20(0.11-0.38)**	**<0.001**	**2.28(1.05-4.94)**	**0.037**
P-xylene (µg/m3)	**0.39(0.19-0.80)**	**0.010**	**0.14(0.06-0.30)**	**<0.001**	1.98(0.86-4.55)	0.106
O-xylene (µg/m3)	0.33(0.19-0.55)	**<0.001**	**0.22(0.13-0.37)**	**<0.001**	**3.03(1.36-6.74)**	**0.006**
M-xylene (µg/m3)	0.59(0.44-0.78)	**<0.001**	**0.43(0.31-0.58)**	**<0.001**	1.39(0.95-2.02)	0.085
TBTEX (µg/m3)	**0.94(0.89-0.99)**	**0.042**	**0.88(0.83-0.94)**	**<0.001**	**1.08(1.01-1.15)**	**0.019**
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	**1.002(1.00-1.003)**	**0.017**	0.99(0.99-1.001)	0.892	**1.003(1.001-1.005)**	**<0.001**

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m^2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, life expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

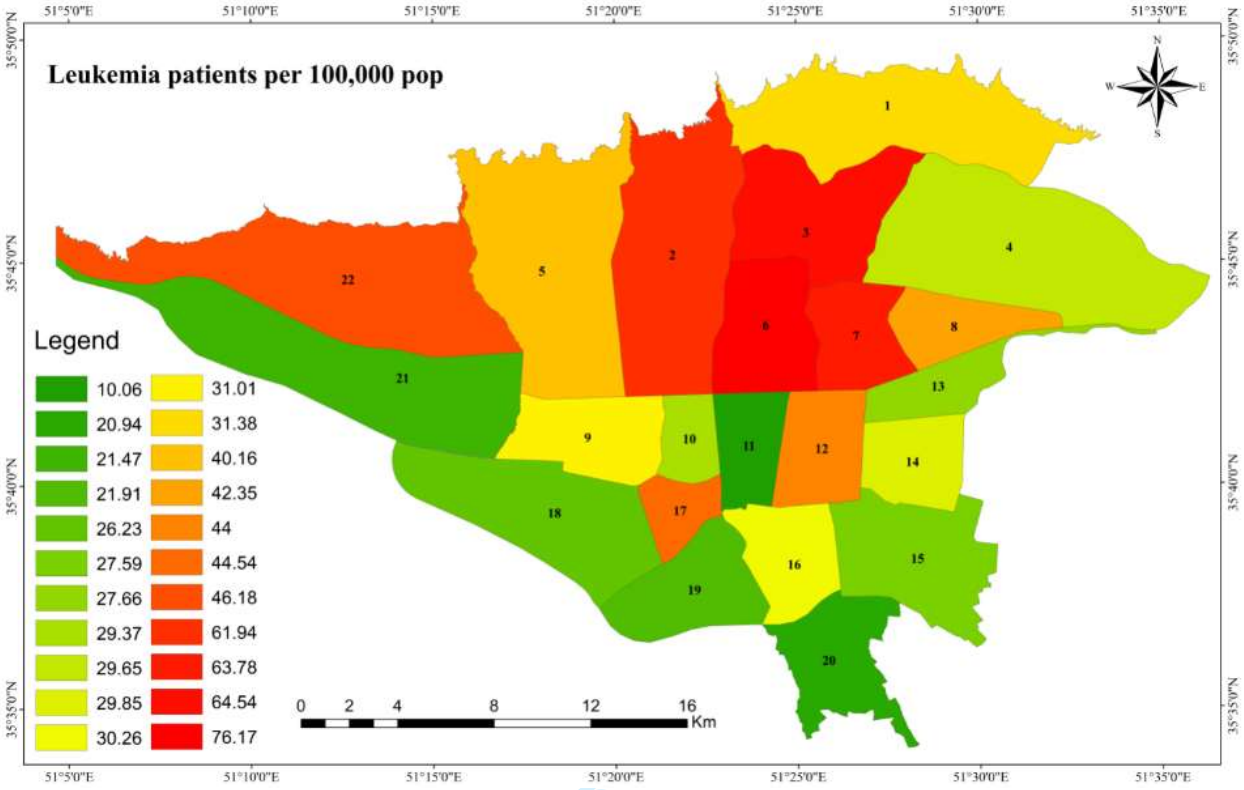

Fig. 1. Spatial distribution of Leukemia cancer patients (number of cases in 100,000) in different areas of Tehran in 2014-2016 (The labels on the map are district numbers).

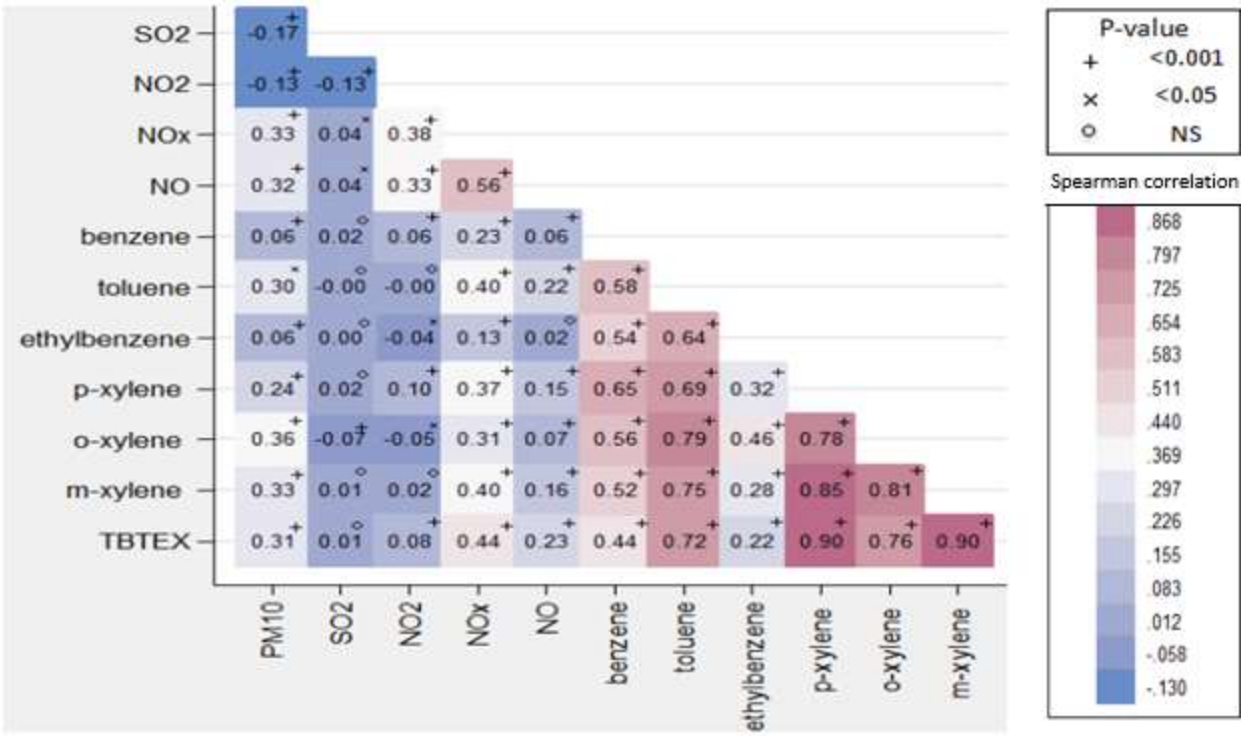

Fig. 2. Spearman correlation matrix of air pollutants in the 22 districts of Tehran, Iran, in 2014-2016.

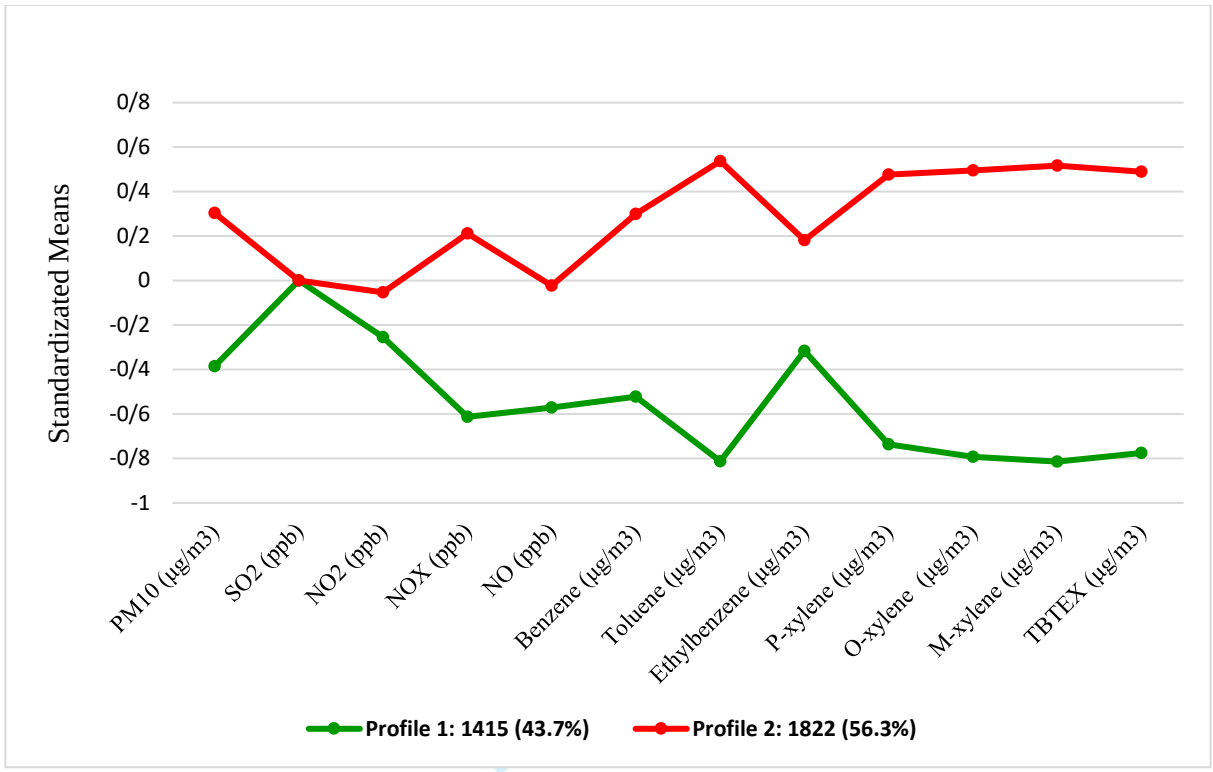

Fig. 3. Standard mean values of pollutants in the two latent profiles in different areas of Tehran, Iran, in 2014-2016.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4,5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4,5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4,5*
Bias	9	Describe any efforts to address potential sources of bias	5
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5,6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5,6
		(*b*) Describe any methods used to examine subgroups and interactions	5,6
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	NA
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	6
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*NA*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	*NA*
		Cross-sectional study—Report numbers of outcome events or summary measures	*6, 16*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	22
		(*b*) Report category boundaries when continuous variables were categorized	16
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	22,23,24
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	23,24
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8,9
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	11

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Multiple Air Pollutants Exposure and Leukemia Incidence in Tehran, Iran from 2010 to 2016: a retrospective cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-060562.R2
Article Type:	Original research
Date Submitted by the Author:	12-May-2022
Complete List of Authors:	khorrami, zahra; Kerman University of Medical Sciences, Epidemiology & Biostatistics Pourkhosravani, Mohsen; Shahid Bahonar University of Kerman Eslahi, Marzieh; Kerman University of Medical Sciences, Department of Epidemiology and Biostatistics Rezapour, Maysam; Mazandaran University of Medical Sciences Akbari, Mohammad Esmail; Shahid Beheshti University of Medical Sciences Amini, Heresh; University of Copenhagen Taghavi-Shahri, Seyed Mahmood; University of Copenhagen, Department of Public Health Künzli, Nino ; Schweizerisches Tropen- und Public Health-Institut Etemad, Koorosh; Shahid Beheshti University of Medical Sciences, Department of Epidemiology Khanjani, Narges ; Shahid Beheshti University of Medical Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health
Keywords:	Leukaemia < HAEMATOLOGY, EPIDEMIOLOGY, Epidemiology < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Multiple Air Pollutants Exposure and Leukemia Incidence in Tehran, Iran from 2010 to 2016: a retrospective cohort study

Zahra Khorrami[1], Mohsen Pourkhosravani[2], Marzieh Eslahi[3], Maysam Rezapour[4], Mohammad Esmail Akbari[5], Heresh Amini[6], Seyed Mahmood Taghavi-Shahri[7], Nino Künzli[8, 9], Koorosh Etemad[10*], Narges Khanjani[11, 12*]

1. Ophthalmic Research Center, Research Institute for Ophthalmology and Vision Science, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Department of Geography and Urban Planning, Shahid Bahonar University of Kerman, Kerman, Iran
3. Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran
4. Amol Faculty of Paramedical Sciences, Mazandaran University of Medical Sciences, Sari, Iran.
5. Cancer Research Center (CRC), Shahid Beheshti University of Medical Sciences, Tehran, Iran.
6. Department of Public Health, University of Copenhagen, Copenhagen, Denmark
7. Section of Environmental Health, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
8. Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland
9. University of Basel, Basel, Switzerland
10. Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran
11. Environmental Health Engineering Research Center, Kerman University of Medical Sciences, Kerman, Iran
12. Monash Centre for Occupational & Environmental Health, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

*Corresponding authors:
1. Prof. Narges Khanjani, Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran.
Tel/Fax: 034-3132-5102 Email: n_khanjani@kmu.ac.ir.
2. Koorosh Etemad, Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
Tel/Fax: 021- 2243-2040 Email: etemadk@gmail.com

Abstract

Objective: Leukemia is one of the most common cancers and may be associated with exposure to environmental carcinogens, especially outdoor air pollutants. The objective of this study was to investigate the association of ambient air pollution and leukemia in Tehran, Iran.

Design: In this retrospective cohort study, data about the residential district of leukemia cases diagnosed from 2010 to 2016 were inquired from the Ministry of Health cancer database. Data from a previous study was used to determine long-term average exposure to different air pollutants in 22 districts of Tehran. Latent profile analysis (LPA) was used to classify pollutants in two exposure profiles. The association between air pollutants and leukemia incidence was analyzed by negative binomial regression.

Setting: Twenty-two districts of Tehran megacity.

Participants: Leukemia patients.

Outcome measures: The outcome variables were incidence rate ratios of Acute Myeloid and Lymphoid Leukemia across the districts of Tehran.

Results:

The districts with higher concentrations for all pollutants were near the city center. The incidence rate ratio (IRR) was positive but non-significant for most of the air pollutants. However, annual mean NOx was directly and significantly associated with total leukemia incidence in the fully adjusted model (IRR (95% CI): 1.03 (1.003, 1.06) per 10 ppb increase). Based on LPA, districts with a higher multiple air-pollutants profile were also associated with higher leukemia incidence (IRR (95% CI): 1.003 (0.99, 1.007) per 1 ppb increase).

Conclusions:

Our study shows that districts with higher air pollution (nitrogen oxides and multi-pollutants) have higher incidence rates of leukemia in Tehran, Iran. This study warrants conducting further research with individual human data and better control of confounding.

Keywords

Air pollution ; multiple pollutants; pollutant profile; Leukemia; Latent profile analysis

Strengths and limitations of this study

- This is the first study for estimating the simultaneous effect of single and multiple ambient air pollutants on the incidence of leukemia in Iran.
- We included potential confounding covariates at area level based on administrative survey data.
- However, control for confounding variables were not done at the individual level.
- We were not able to adjust for human relocation or migration.

Background

Leukemia is one of the most common hematological malignancies among adults and children [1]. In 2020, leukemia was amongst the 15 more common cancers worldwide and contributed to 2.5 % of all new cases of cancer, and 3.1% of all cancer death. Worldwide, the age standardized incidence and mortality rates of leukemia are 6.3 per 100,000 and 4.0 per 100,000 for men and 4.5 per 100,000 and 2.7 per 100,000 in women, respectively. In countries with low and medium human development index (HDI), the age- standardized incidence and mortality rates of leukemia are, respectively, 3.8 per 100,000 and 3.0 per 100,000 for men, and 2.9 per 100,000 and 2.0 per 100,000 for women [2]. In Iran where cancer is the third leading cause of death after cardiovascular disease, mortality and traffic accidents, leukemia is the 6th most common cancer in men with age standardized rates of 8.32 per 100,000, and the 5th in women with age standardized rates of 5.40 per 100,000 [3].

Outdoor air pollution may have serious adverse effects on human health, including cancer [4]. It has been proven that some environmental factors including ambient air pollution are carcinogenic to humans [5].

The etiology of leukemia remains mostly unknown. Some studies have reported that adverse gene-environment interactions may lead to leukemia [1]. A recent meta-analysis on the association between outdoor air pollution and childhood leukemia showed a positive linear association between benzene exposure and risk of childhood leukemia, particularly for acute myeloid leukemia, among children under 6 years of age, and nitrogen dioxide at high levels was associated with leukemia [6].

In the current study, we aimed to examine the association between estimates of district level averages of multiple ambient air pollutants and acute leukemia incidence across Tehran districts using latent profile analysis (LPA) method. LPA is a robust technique, mainly used to identify subtypes of homogeneous latent classes or subgroups within a large heterogeneous group. This iterative process, clusters similar profiles together to generate distinct subgroups/classes [7].

Methodology

Research location

The present retrospective cohort study was carried out based on annual mean air pollution levels in 22 districts of Tehran megacity, which is the capital of Iran. According to the World Population Review report, Tehran's 2021 population was estimated to be 9.2 million inhabitants. Tehran includes 22 districts in which each include from 174,239 to 919,001 residents according to the latest 2016 census [8]. Population density is higher in the central, western, and southern regions [9]. Tehran suffers from severe ambient air pollution as documented by numerous studies [9-12]. The central districts with higher population densities (five districts of 2, 6, 10, 11, and 12 in central Tehran) have more air pollution [9].

Data Sources

Leukemia data

Information about leukemia patients (acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) residing in Tehran, diagnosed between 2010 and 2016 and their residential address (on a district basis) were obtained from the Ministry of Health's Cancer Registry.

Exposure assessment

The annual mean concentrations of several pollutants in the 22 districts of Tehran were obtained from land use regression (LUR) models developed in previous studies, for PM_{10}, SO_2, NO, NO_2, and NO_X in Tehran, which were based on measurements conducted at 23 regulatory network monitoring sites in 2010 [10, 11]. Data about volatile organic compounds (VOCs) were taken from a previous study [12]. The average of sampling site estimates for each pollutant, in each district was determined and included in the analyzes of this study.

Covariates

District level data including urban green space per capita, life expectancy, and socioeconomic status, was extracted from the Urban Health Equity Assessment and Response Tool (Urban HEART-2), which has been conducted in the 22 districts of Tehran. A detailed description of the Urban HEART-2 study can be found elsewhere [13]**.** The socio-economic indicators of the 22 districts of Tehran was extracted from a study conducted by Sadeghi et al [14].

Statistical analyses

This study used latent profile analysis (LPA) to investigate the patterns of multiple air pollutants exposure. LPA is a statistical method to identify unobserved subgroups (profile) within

populations based on observed variables. LPA has several advantages over traditional methods, such as cluster analysis, e.g. it does not require researchers to specify the number of profiles in advance, which is more suitable for addressing research questions that are exploratory in nature. Moreover, unlike cluster analysis that assigns individuals to clusters absolutely, LPA assigns individuals to subgroups probabilistically [15].

Latent profile analysis (LPA) as a person-centered approach can be used to examine the patterns of multiple air pollutants. LPA is a statistical method for identifying unobserved subgroups within populations based on observed indicators. In contrast to traditional methods, such as cluster analysis, LPA has several advantages. LPA does not require researchers to determine the number of profiles beforehand, and this is more suitable to answer research questions that are exploratory in nature. Also, empirical indicators are available to determine the optimal number of profiles. In addition, LPA allocates individuals to subgroups probabilistically, taking into account the rate of classification uncertainty, and uses multiple statistical indices for determining the optimal number of subgroups.

Most air pollutants in this study had a skewed distribution and were transformed by natural logarithm before LPA, except PM_{10} and SO_2. Several latent profile models were performed, ranging from two to five latent profiles. The most appropriate number of subgroups was identified based on statistical criteria and profile interpretability. The statistical criteria included the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample-size adjusted Bayesian information criterion (a-BIC), the Lo-Mendell-Rubin likelihood ratio test (LMR-LRT), and the Vuong-Lo-Mendell-Rubin likelihood ratio tests (VLMR-LRT). Smaller values of the AIC, BIC, and aBIC indicate a better fit model [16].

A significant p-value of the LMR LRT and VLMR LRT (i.e. $P < 0.05$) indicates a significant improvement in model fit in the k-class model compared to the (k - 1)-class model and thus rejects the (k - 1)-class model and suggests choosing a model with k classes.

In other words, each number of classes (or profiles, shown as k) is compared to the number of classes which is 1 unit less (k-1), and if the VLRT and LMRT values are not significant for higher classes, the model with less classes (k-1) is preferred and will get chosen [17].

Data preparation was done in Stata version 14. LPA was conducted using Mplus version 7.4. One-way ANOVA and post hoc follow-up tests were used to investigate the differences between

profiles of multiple pollutions in terms of each component of air pollution. Management of missing data and other statistical preparation details have been mentioned in our previous publication [8].

Kolmogorov Smirnov test was used to test the normality of the pollutants data and because the data were not normally distributed, Spearman's correlation test was used to estimate the correlation between pollutants. As the number of leukemia 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, and district level (by using them as a covariate). Statistical analyses were performed using Mplus version 7.4, and Stata version 14 (Stata Corp LLC; College Station, TX, USA). ArcGIS version 10.8 was used for exposure assignment to districts and data visualization.

Patient and public involvement

Patients and the public were not involved in the design and conduct of this research.

Results

Summary statistics of patients, air pollutants, and information about the area under study are shown in Table 1. The total number of leukemia cases in 2010–2016 in all districts of Tehran was 3,237. The distribution of leukemia patients in different districts of Tehran is shown in Figure 1. The highest numbers of leukemia patients per 100,000 people were in districts 6, 3, 7 and 2, respectively (Figure 1).

As shown in Figure 2, there was a strong correlation between pollutants, most notably for VOCs (benzene, toluene, m-xylene, p-xylene, o-xylene and TBTEX) (p-value<0.001). The positive correlation between NO and NOx was weaker (r=0.56).

Fit indices for the different LPA models are displayed in Table 2. Several latent profile models were considered. Although, the AIC, BIC, and aBIC of the 2-profile were more than other models, the LMR of LRT (Lo–Mendell–Rubin likelihood ratio test) and VLMR (Vuong-Lo-Mendell-Rubin likelihood ratio test) were significant in the two-profile model.

Figure 3 shows the multi-pollution profiles. Profile 1 had the lowest scores for all pollutants including (PM_{10}, SO2, NO, NO_2, NO_X, benzene, toluene, ethylbenzene, m-xylene, p-xylene, o-xylene and TBTEX). We labeled this profile as "low 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 two profiles, except for SO_2.

Table 4 shows the IRR estimates and 95% confidence interval (CI) by single-pollutant and multiple-pollutant multivariable negative binomial regression models, adjusted for age, gender, socioeconomic status and life expectancy.

In single-pollutant models, NO_2 and NO_X were significantly associated with increased leukemia incidence in model 1 (adjusted for age and gender), and also in model 2 (adjusted in addition for urban green space per capita) with IRR (95% CI) of model 2, respectively, 1.35 (1.11-1.64) and 1.07 (1.03-1.11) per 10 ppb increase in pollutants. In model 3, after adjustment for age, gender, socioeconomic status, and life expectancy, only NO_X was significantly associated with increased leukemia incidence with an IRR (95% CI) of 1.03 (1.003-1.06) per 10 ppb increase in NOx. However, NO_2 was borderline significantly associated with increased leukemia incidence (IRR=1.19, CI 95%=0.99-1.43) in this model.

In multi-pollutant models, the high multiple-air-pollutants profile was associated with higher leukemia incidence when compared with the low multiple-air-pollutants profile, but this association did not reach significance.

Tables 5 and 6 show the IRR (95% CI) estimates by single-pollutant and multiple-pollutant multivariable negative binomial regression models for acute myeloid and lymphoid leukemia incidence, respectively. NO, NO_2 and NO_x were related to acute increased lymphoid leukemia incidence, while NO_2 and TBTEX were related to increased acute myeloid leukemia incidence.

Tables 7 and 8 show the IRR (95% CI) estimates by single-pollutant and multiple-pollutant multivariable negative binomial regression models for total acute leukemia incidence, respectively in children and adults. Increase in all single pollutants except SO_2 and high multiple pollutants were related to increased acute leukemia incidence in children (≤14 years old), while nitrogen oxides, toluene, ethylbenzene, o-xylene and TBTEX were related to increased acute leukemia incidence among adults.

Discussion

The findings of our study show that exposure to air pollutants, such as nitrogen oxides and VOCs may be associated with the incidence of leukemia. Our study was the first to investigate the effect of single and multiple ambient air pollutants on leukemia in Iran.

Leukemia is one of the most common cancers in children and adults. The cause of leukemia is currently unknown [18]. However, some sources have suggested that genetic and transgenic mutations due to environmental factors may contribute to leukemia [1, 19]. A study in Shanghai, China showed that air pollution from industrial waste gas emissions was associated with the incidence of several cancers including leukemia [20].

Many experimental studies have shown that air pollutants may be carcinogenic by causing DNA damage and mutations [21-26].

A review indicated that exposure to air pollutants is associated with leukemia stronger than other cancers [18]. Carlos-Wallace et al. also conducted a meta-analysis and reported associations between childhood leukemia and benzene exposure. They indicated that in studies that evaluated benzene versus all solvents, the pooled RR (Relative Risk from meta-analysis) was higher for maternal exposure compared to paternal exposure. For NO_2, an excess risk was reported in concentration-response meta-analysis from 40 $\mu g/m^3$ to 60 $\mu g/m^3$; however, the increase was not statistically significant and was mainly related to ALL [6].

Our study showed some associations between benzene and leukemia incidence among children. Similarly, a study in the United Kingdom revealed that there was an increased risk of leukemia from low-level exposure to benzene from smoking, and that benzene may contribute to up to a third of smoking-induced leukemia. The contribution of benzene to smoking-induced leukemia was estimated to be between 9% and 24%. For AML this contribution was estimated as 11-30% [27]. Janitz et al. conducted a case-control study between 1997 and 2012 including 307 patients with acute leukemia from the Oklahoma Central Cancer Registry and 1013 controls matched on week of birth from birth certificates and showed no association between benzene and childhood leukemia, in Oklahoma. However, based on the analysis of quartiles they indicated that there was a positive concentration-response relation between benzene and AML [28].

Our study indicated that NO_2 increased the incidence of acute lymphoid leukemia, and total leukemia among children and adults. Similarly, findings of the population-based study by

Ribeiro et al in São Paulo, Brazil has shown that NO_2 and traffic density were associated with Hodgkin lymphoma and lymphoid leukemia in children; and the variations in the incidence rate ratios in gender and SES groups may be because of differences in underlying risk and exposure profiles [29]. In addition, Raaschou-Nielsen et al. conducted a nationwide case-control study in Denmark and indicated that long-term exposure to traffic-related air pollutants (NO_x and NO_2) was associated with acute myeloid leukemia, but not other subtypes of leukemia, in the general population [30]. Our study revealed that NO, NO_2 and NO_x are related to increased acute lymphoid leukemia incidence, while NO_2 and TBTEX are related to increased acute myeloid leukemia incidence. While, a Canadian population-based case–control study revealed a weak association between all forms of leukemia only at low concentrations of NO_2. The study showed an 'n-shaped' response function between exposure to NO_2 and all forms of leukemia. The OR was 1.20 (95% CI: 0.97-1.48) from the 10^{th} percentile (4.51 ppb) to the median (14.66 ppb), then the OR decreased to 0.79 (95% CI: 0.68, 0.93) from the 75th percentile (22.75) to the 90th (29.7 ppb) range [31]. Some differences in results may be attributable to differences in settings and population characteristics.

In a large nationwide population-based case-control study in Denmark, Taj et al. showed a high risk for overall leukemia in association with secondary inorganic aerosols including nitrate (NO_3), and that AML was associated with NO_3 [32], which was similar to our findings.

Studies focusing on the association between childhood leukemia and exposure to air pollutants have had inconsistent results. Ghosh et al in California revealed that prenatal exposure to traffic-related air pollution was associated with the risk of ALL [33]. A population-based study from Canada suggested that exposure to ambient air pollution during the first trimester of pregnancy, may cause astrocytoma and ALL [34]. However, Peckham-Gregory et al. conducted a population-based case-control study in Texas to evaluate the association between maternal residential proximity to major roadways and developing ALL and AML in children, and reached different results. They indicated that mothers who lived closer than 500 meters to a major roadway and mothers who lived in high roadway density areas were not more likely to have a child with ALL or AML [35].

Although many studies have shown a possible relation between air pollution and leukemia, there is still a need for more high-quality studies with higher sample sizes, and better control of confounders.

Strengths and limitation

In this study, we estimated the simultaneous effect of several different air pollutants on the incidence of leukemia. However, our study had several limitations. We did not use individual-level data; thus, we were not able to control for confounding variables at individual-level, such as blood group, family history of cancer, taking medicine during pregnancy, parents' job, history of radiation, smoking, and other factors such as genetics, nutrition status, cultural context, and behavioral patterns. But we did include potential confounding covariates at regional level.

The other major limitation of this study is our estimates of exposure. First, our study used some more recent air pollution data, not past life-time exposure estimates, because we did not have a better option. In addition, we did not have data on the length of residence of patients in the study regions, and life-time relocation or migration. Second, we assigned the same exposure concentrations to all people living in the same area. Although the mean area level values may well reflect the average exposure levels of the inhabitants, our approach ignored within-area variation, which has been demonstrated in previous studies in Tehran [10-12].

Conclusion

This is the first study to examine the association between multiple air pollutants and leukemia incidence in Iran. Our findings suggest that exposure to VOCs, nitrogen oxides and/or multiple ambient air pollutants may be associated with increased leukemia incidence in Tehran. Further research with individual data and better control of confounding covariates is needed to confirm the role of air pollution in human leukemia.

List of abbreviations

LPA: Latent profile analysis; IRR: Incidence Rate Ratios; HDI: Human Development Index; ASR: Age Standardized Rates; LUR: Land Use Regression; VOCs: Volatile Organic

Compounds; AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion; LMR-LRT: Lo-Mendell-Rubin Likelihood Ratio Test; CI: Confidence Intervals; NB: Negative Binomial; ALL: Acute Lymphoblastic Leukemia; AML: Acute Myeloid Leukemia.

Declarations

Acknowledgements

The authors thank the Cancer Department of the Iranian Ministry of Health, and the Tehran Air Quality Control Company (AQCC), which provided that data for this project. HA is supported by Novo Nordisk Foundation Challenge Program: Harnessing the Power of Big Data to Address the Societal Challenge of Aging (NNF17OC0027812).

Authors' contributions

Z.K., N.K. and M.R study design and Z.K., N.K., M.P conceptualization and M.R, Z.K. and M.P data cleaning and K.E., M.E.A, H.A., S.M.T.-S., N.K data management and statistical analysis and statistical inference and Z.K., M.R., M.P. and H.A Z.K., M.E, N.K. and M.R writing the original draft and Z.K., N.K., M.P., M.R., M.E.A, K.E., S.M.T.-S., N.K., H.A review and editing. All authors participated in revision of the final draft and agreed on the final content.

Funding

This study was financially supported by the Cancer Research Center of Shahid Beheshti University of Medical Sciences, Tehran by Grant No. 25544. HA is supported by Novo Nordisk Foundation Challenge Programme: Harnessing the Power of Big Data to Address the Societal Challenge of Aging [NNF17OC0027812].

Competing interests

The authors declare no competing interests.

Patient consent for publication

Not required

Availability of data and materials

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.

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. As the data was inquired in aggregated form and anonymously, informed consent from individuals or their family was not required. This project was approved by the Ethics in Research Committee of Shahid Beheshti University of Medical Sciences. Ethics Code: IR.SBMU.CRC.REC.1400.003.

Consent for publication

Not applicable.

Fig. 1. Spatial distribution of Leukemia patients (number of cases in 100,000) in different areas of Tehran in 2014-2016 (The labels on the map are district numbers).

Fig. 2. Spearman correlation matrix of air pollutants in the 22 districts of Tehran, Iran, in 2014-2016.

Fig. 3. Standard mean values of pollutants in the two latent profiles in different areas of Tehran, Iran, in 2014-2016.

References:

1. Lien, S.A., et al., *Meta-prediction of MTHFR gene polymorphism-mutations, air pollution, and risks of leukemia among world populations.* Oncotarget, 2017. **8**(3): p. 4387-4398.
2. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021.
3. Roshandel, G., et al., *Cancer incidence in Iran in 2014: results of the Iranian National Population-based Cancer Registry.* Cancer epidemiology, 2019. **61**: p. 50-58.
4. Turner, M.C., et al., *Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations.* CA Cancer J Clin, 2020.
5. Ou, J.Y., A.C. Kirchhoff, and H.A. Hanson, *Air Pollution across the Cancer Continuum: Extending Our Understanding of the Relationship between Environmental Exposures and Cancer.* Cancer Epidemiol Biomarkers Prev, 2020. **29**(10): p. 1876-1879.
6. Filippini, T., et al., *Association between Outdoor Air Pollution and Childhood Leukemia: A Systematic Review and Dose-Response Meta-Analysis.* Environ Health Perspect, 2019. **127**(4): p. 46002.
7. Fernández, R.S., et al., *Psychological distress associated with COVID-19 quarantine: Latent profile analysis, outcome prediction and mediation analysis.* J Affect Disord, 2020. **277**: p. 75-84.
8. Khorrami, Z., et al., *Multiple air pollutant exposure and lung cancer in Tehran, Iran.* Scientific Reports, 2021. **11**(1): p. 1-11.
9. Ghaedrahmati, S. and M. Alian, *Health risk assessment of relationship between air pollutants' density and population density in Tehran, Iran.* Human and Ecological Risk Assessment: An International Journal, 2019. **25**(7): p. 1853-1869.

10. 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.* Science of the total environment, 2014. **488**: p. 343-353.
11. Amini, H., et al., *Annual and seasonal spatial models for nitrogen oxides in Tehran, Iran.* Scientific reports, 2016. **6**(1): p. 1-11.
12. Amini, H., et al., *Land use regression models for Alkylbenzenes in a middle eastern megacity: Tehran study of exposure prediction for environmental Health Research (Tehran SEPEHR).* Environmental science & technology, 2017. **51**(15): p. 8481-8490.
13. 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.* Medical journal of the Islamic Republic of Iran, 2013. **27**(4): p. 236.
14. Sadeghi, R. and N. Zanjari, *The inequality of development in the 22 districts of Tehran metropolis.* Social Welfare Quarterly, 2017. **17**(66): p. 149-184.
15. Mindrila, D.L., *A typology of child school behavior: Investigation using latent profile analysis and cluster analysis.* Psychology in the Schools, 2016. **53**(5): p. 471-487.
16. Nylund, K.L., T. Asparouhov, and B.O. Muthén, *Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study.* Structural equation modeling: A multidisciplinary Journal, 2007. **14**(4): p. 535-569.
17. Tein, J.-Y., S. Coxe, and H. Cham, *Statistical power to detect the correct number of classes in latent profile analysis.* Structural equation modeling: a multidisciplinary journal, 2013. **20**(4): p. 640-657.
18. Pourvakhshoori, N., et al., *The association between air pollution and cancers: controversial evidence of a systematic review.* Environ Sci Pollut Res Int, 2020. **27**(31): p. 38491-38500.
19. Gao, Y., et al., *Quantitative assessments of indoor air pollution and the risk of childhood acute leukemia in Shanghai.* Environ Pollut, 2014. **187**: p. 81-9.
20. Cong, X., *Air pollution from industrial waste gas emissions is associated with cancer incidences in Shanghai, China.* Environ Sci Pollut Res Int, 2018. **25**(13): p. 13067-13078.
21. Smith, M.T., *Advances in understanding benzene health effects and susceptibility.* Annual review of public health, 2010. **31**: p. 133-148.
22. Whysner, J., et al., *Genotoxicity of benzene and its metabolites.* Mutation Research/Reviews in Mutation Research, 2004. **566**(2): p. 99-130.
23. Mondrala, S. and D.A. Eastmond, *Topoisomerase II inhibition by the bioactivated benzene metabolite hydroquinone involves multiple mechanisms.* Chemico-biological interactions, 2010. **184**(1-2): p. 259-268.
24. Yang, J., et al., *PTEN methylation involved in benzene-induced hematotoxicity.* Experimental and molecular pathology, 2014. **96**(3): p. 300-306.
25. Koehler, C., et al., *Nitrogen dioxide is genotoxic in urban concentrations.* Inhalation toxicology, 2013. **25**(6): p. 341-347.
26. Kampa, M. and E. Castanas, *Human health effects of air pollution.* Environmental pollution, 2008. **151**(2): p. 362-367.
27. Fiebelkorn, S. and C. Meredith, *Estimation of the leukemia risk in human populations exposed to benzene from tobacco smoke using epidemiological data.* Risk Analysis, 2018. **38**(7): p. 1490-1501.
28. Janitz, A.E., et al., *Benzene and childhood acute leukemia in Oklahoma.* Environ Res, 2017. **158**: p. 167-173.

29. Ribeiro, A.G., et al., *Residential traffic exposure and lymphohematopoietic malignancies among children in the city of São Paulo, Brazil: An ecological study.* Cancer Epidemiol, 2021. **70**: p. 101859.
30. Raaschou-Nielsen, O., et al., *Traffic-related air pollution and risk for leukaemia of an adult population.* Int J Cancer, 2016. **138**(5): p. 1111-7.
31. Winters, N., et al., *Exposure to ambient air pollution in Canada and the risk of adult leukemia.* Science of the Total Environment, 2015. **526**: p. 153-176.
32. Taj, T., et al., *Exposure to PM(2.5) constituents and risk of adult leukemia in Denmark: A population-based case-control study.* Environ Res, 2020: p. 110418.
33. Ghosh, J.K., et al., *Prenatal exposure to traffic-related air pollution and risk of early childhood cancers.* Am J Epidemiol, 2013. **178**(8): p. 1233-9.
34. Lavigne, É., et al., *Maternal exposure to ambient air pollution and risk of early childhood cancers: A population-based study in Ontario, Canada.* Environ Int, 2017. **100**: p. 139-147.
35. Peckham-Gregory, E.C., et al., *Maternal Residential Proximity to Major Roadways and the Risk of Childhood Acute Leukemia: A Population-Based Case-Control Study in Texas, 1995-2011.* Int J Environ Res Public Health, 2019. **16**(11).

Table 1. Description of characteristics of patients, air pollution, and district-level covariates.

Individual-Level variables	
Number of Leukemia patients	3,237
Age of patients (year)	
Median (1st-3st quartile range)	55 (16-69.1)
Gender of patients (male)	
Percent (number)	59.7 (1934)
Grading	
I	4.8(156)
II	33.2(1076)
III	2.8(90)
IV	1(32)
Unknown	58.2(1883)
Topography	
Lymphoid	68.9(2231)
Myeloid	31.1(1006)
Air pollutants	Median (1^{st}-3^{st} quartile range)
PM_{10} ($\mu g/m^3$)	101.32(82.35-123.82)
SO_2 (ppb)	52.42(26.38-77.19)
NO_2 (ppb)	46.43(39.95-55.74)
NO (ppb)	78.13(44.47-118.84)
NO_X (ppb)	112.12(83.24-158.49)
benzene ($\mu g/m^3$)	8.12(7.02-9.85)
toluene ($\mu g/m^3$)	24.96(20.85-29.45)
ethylbenzene ($\mu g/m^3$)	5.90(4.97-6.94)
p-xylene ($\mu g/m^3$)	5.71(4.88-6.52)
o-xylene ($\mu g/m^3$)	5.84(4.86-7.46)
m-xylene ($\mu g/m^3$)	10.71(9.06-12.81)
TBTEX ($\mu g/m^3$)	60.57(52.29-70.15)
District level variables	Median (1^{st}-3^{st} quartile range)
Number of districts	22
Total population (persons)	319,443 (263,536-421,225)
Urban green space, per capita (m^2 per 1000 people)	13.9 (10.7-18)
Life expectancy (year)	75.9 (75.4-76.7)
Socioeconomic status *	49.95 (44.21-53.34)

* 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.

Table 2. Fit indices for different latent profile models with number of profiles ranging from 2 to 5.

Profile	Log-likelihood	AIC	BIC	aBIC	LMR LRT	VLMR LRT	entropy
2 profile	**-35505.2**	**71084.5**	**71309.6**	**71192**	**13725.8***	**13856.4***	**0.908**
3 profile	-31630.4	63360.8	63665	63506.1	7676.6	7749.6	0.941
4 profile	-29593.4	59312.9	59696.1	59495.9	4035.5	4073.9	0.951
5 profile	-27590.1	55332.2	55794.5	55553	3968.9	4006.7	0.968

Note: AIC Akaike's Information Criterion, BIC Bayesian Information Criterion, aBIC Sample Size Adjusted BIC, and LMR, Lo–Mendell–Rubin likelihood ratio test; VLMR, Vuong-Lo-Mendell-Rubin likelihood ratio test *P value <0.001.

Models with a significant Lo–Mendell–Rubin likelihood ratio test and Vuong-Lo-Mendell-Rubin likelihood ratio test were preferred. The bolded fit indices indicate that profile 2 was the optimal latent profile model.

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

pollutant	profile	Mean	SD	T	P-value
PM10 (µg/m3)	profile1	87.2	38.3	-21.6	<0.001
	profile2	114.2	31.1		
SO2 (ppb)	profile1	54.8	45.7	-0.32	0.74
	profile2	55.3	53.7		
NO2 (ppb)	profile1	48.4	18.2	-3.84	<0.001
	profile2	50.7	15.7		
NO (ppb)	profile1	68.9	51.6	-19.8	<0.001
	profile2	108.9	63.5		
NOX (ppb)	profile1	91.6	38.8	-33.55	<0.001
	profile2	163.8	80.7		
Benzene (µg/m3)	profile1	8.9	3.2	-26.3	<0.001
	profile2	10.1	3.6		
Toluene (µg/m3)	profile1	19.6	3.5	-56.9	<0.001
	profile2	30.2	6.8		
Ethylbenzene (µg/m3)	profile1	4.7	1.1	-17.4	<0.001
	profile2	7.3	6.1		
P-xylene (µg/m3)	profile1	4.7	0.6	-49.9	<0.001
	profile2	6.7	1.5		
O-xylene (µg/m3)	profile1	4.8	0.8	-53.8	<0.001
	profile2	7.4	1.7		
M-xylene (µg/m3)	profile1	8.6	1.5	-57.1	<0.001
	profile2	13.2	2.9		
TBTEX (µg/m3)	profile1	49.5	7.8	-53.1	<0.001
	profile2	72.9	16.5		

Table 4. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on total Leukemia incidence across the districts of Tehran.

	Model 1		Model 2		Model 3	
	IRR (95% CI)	P-value	IRR (95% CI)	P-value	IRR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM_{10} (µg/m3)	0.95(0.87-1.05)	0.352	0.92(0.82-1.03)	0.165	1.03(0.95-1.12)	0.376
Annual SO_2 (ppb)	0.99(0.89-1.09)	0.901	0.99(0.98-1.01)	0.775	0.97(0.91-1.05)	0.575
Annual NO_2 (ppb)	**1.35(1.11-1.64)**	**0.002**	**1.35(1.11-1.64)**	**0.002**	1.19(0.99-1.43)	0.062
Annual NO (ppb)	1.02(0.98-1.07)	0.238	1.03(0.97-1.08)	0.235	1.02(0.99-1.06)	0.130
Annual NO_X (ppb)	**1.04(1.01-1.08)**	**0.003**	**1.07(1.03-1.11)**	**<0.001**	**1.03(1.003-1.06)**	**0.029**
Benzene (µg/m3)	0.97(0.85-1.11)	0.696	0.97(0.85-1.11)	0.965	0.92(0.83-1.02)	0.119
Toluene (µg/m3)	1.11(0.78-1.56)	0.550	1.14(0.71-1.85)	0.570	1.11(0.85-1.44)	0.436
Ethylbenzene (µg/m3)	1.08(0.27-4.28)	0.906	0.97(0.19-4.91)	0.974	1.09(0.38-3.16)	0.865
P-xylene (µg/m3)	1.23(0.14-10.23)	0.844	1.002(0.04-22.77)	0.999	1.24(0.23-6.42)	0.797
O-xylene (µg/m3)	0.48(0.09-2.57)	0.397	0.29(0.04-2.03)	0.214	0.77(0.21-2.84)	0.705
M-xylene (µg/m3)	1.14(0.48-2.72)	0.760	1.11(0.28-4.36)	0.876	1.20(0.61-2.34)	0.589
TBTEX (µg/m3)	1.04(0.89-1.23)	0.561	1.07(0.84-1.36)	0.575	1.03(0.91-1.17)	0.564
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.003(0.99-1.009)	0.227	1.005(0.99-1.01)	0.168	1.003(0.99-1.007)	0.168

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m^2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, life expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 5. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on Acute Myeloid Leukemia incidence in the districts of Tehran.

	Model 1		Model 2		Model 3	
	RR (95% CI)	P-value	RR (95% CI)	P-value	RR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM10 (µg/m3)	0.97(0.88-1.06)	0.547	0.94(0.82-1.08)	0.542	1.01(0.89-1.14)	0.846
Annual SO2 (ppb)	1.004(0.88-1.13)	0.949	1.008(0.88-1.15)	0.904	1.01(0.91-1.11)	0.814
Annual NO2 (ppb)	**1.35(1.11-1.63)**	**0.002**	**1.35(1.11-1.63)**	**0.002**	1.22(0.98-1.52)	0.068
Annual NO (ppb)	1.02(0.98-1.07)	0.260	1.03(0.98-1.09)	0.172	0.99(0.94-1.05)	0.966
Annual NOX (ppb)	0.96(0.93-1.001)	0.057	**0.95(0.92-0.99)**	**0.035**	0.97(0.939-1.01)	0.240
Benzene (µg/m3)	1.71(0.44-6.66)	0.436	4.93(0.60-40.17)	0.136	1.46(0.38-5.47)	0.574
Toluene (µg/m3)	1.19(0.83-1.71)	0.334	1.71(0.97-3.02)	0.061	1.23(0.85-1.77)	0.254
Ethylbenzene (µg/m3)	1.52(0.27-8.53)	0.634	4.59(0.25-82.81)	0.301	1.68(0.31-9.03)	0.545
P-xylene (µg/m3)	1.08(0.87-1.35)	0.454	23.75(0.55-102.3)	0.099	1.78(0.18-17.66)	0.620
O-xylene (µg/m3)	0.81(0.15-4.31)	0.804	0.77(0.06-8.66)	0.837	0.71(0.13-3.87)	0.701
M-xylene (µg/m3)	1.63(0.61-4.31)	0.324	3.03(0.78-11.69)	0.107	1.83(0.74-4.52)	0.190
TBTEX (µg/m3)	1.11(0.92-1.31)	0.255	**1.41(1.07-1.85)**	**0.014**	1.09(0.91-1.32)	0.314
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.003(0.99-1.01)	0.265	1.003(0.99-1.01)	0.259	1.004(0.99-1.01)	0.113

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, Life Expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 6. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on Acute Lymphoid Leukemia incidence in the districts of Tehran.

	Model 1		Model 2		Model 3	
	RR (95% CI)	P-value	RR (95% CI)	P-value	RR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM10 (µg/m3)	0.94(0.87-1.02)	0.207	0.91(0.83-1.01)	0.096	0.99(0.91-1.07)	0.822
Annual SO2 (ppb)	1.08(0.96-1.21)	0.181	1.08(0.96-1.22)	0.172	1.03(0.95-1.12)	0.428
Annual NO2 (ppb)	**1.29(1.07-1.55)**	**0.006**	**1.29(1.07-1.54)**	**0.006**	1.07(0.90-1.27)	0.425
Annual NO (ppb)	1.03(0.98-1.08)	0.157	1.04(0.99-1.10)	0.101	**1.04(0.99-1.08)**	**0.052**
Annual NOX (ppb)	**1.04(1.01-1.08)**	**0.010**	**1.06(1.02-1.10)**	**0.002**	1.01(0.98-1.05)	0.310
Benzene (µg/m3)	1.21(0.36-4.01)	0.754	1.33(0.27-6.48)	0.722	0.46(0.17-1.24)	0.129
Toluene (µg/m3)	1.01(0.72-1.41)	0.949	1.006(0.65-1.54)	0.978	0.92(0.71-1.21)	0.585
Ethylbenzene (µg/m3)	0.84(0.33-2.13)	0.723	0.78(0.27-2.24)	0.658	0.69(0.32-1.46)	0.336
P-xylene (µg/m3)	0.41(0.04-3.51)	0.422	0.21(0.01-3.13)	0.258	0.28(0.05-1.42)	0.126
O-xylene (µg/m3)	0.36(0.08-1.57)	0.175	0.23(0.04-1.26)	0.091	0.32(0.10-1.03)	0.058
M-xylene (µg/m3)	0.83(0.31-2.16)	0.705	0.64(0.16-2.55)	0.533	0.58(0.28-1.20)	0.144
TBTEX (µg/m3)	1.001(0.85-1.17)	0.988	0.99(0.80-1.23)	0.971	0.95(0.84-1.08)	0.433
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	1.002(0.99-1.009)	0.427	1.003(0.99-1.01)	0.343	0.99(0.99-1.004)	0.884

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space, per capita (m2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status Life Expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 7. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on total Leukemia incidence in children (≤14 years old) across the districts of Tehran.

	Model 1		Model 2		Model 3	
	IRR (95% CI)	P-value	IRR (95% CI)	P-value	IRR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM_{10} (µg/m3)	**0.87(0.84-0.92)**	**<0.001**	**0.78(0.74-0.82)**	**<0.001**	**1.11(1.03-1.18)**	**0.003**
Annual SO_2 (ppb)	**0.95(0.91-0.99)**	**0.036**	0.96(0.92-1.008)	0.106	0.99(0.95-1.04)	0.712
Annual NO_2 (ppb)	**1.45(1.32-1.61)**	**<0.001**	**1.45(1.31-1.60)**	**<0.001**	**1.21(1.08-1.35)**	**0.001**
Annual NO (ppb)	**0.93(0.91-0.96)**	**<0.001**	**0.90(0.88-0.93)**	**<0.001**	**1.05(1.02-1.09)**	**0.005**
Annual NO_X (ppb)	**1.04(1.02-1.07)**	**<0.001**	**1.04(1.01-1.07)**	**0.004**	**1.08(1.05-1.10)**	**<0.001**
Benzene (µg/m3)	0.49(0.23-1.05)	0.068	**0.30(0.14-0.67)**	**0.003**	**5.87(2.43-14.16)**	**<0.001**
Toluene (µg/m3)	**0.75(0.63-0.89)**	**0.001**	**0.63(0.52-0.76)**	**<0.001**	**1.77(1.42-2.22)**	**<0.001**
Ethylbenzene (µg/m3)	**0.17(0.07-0.42)**	**<0.001**	**0.06(0.02-0.16)**	**<0.001**	**21.78(6.02-78.85)**	**<0.001**
P-xylene (µg/m3)	**0.14(0.05-0.43)**	**0.001**	**0.02(0.01-0.08)**	**<0.001**	**17.83(4.22-75.29)**	**<0.001**
O-xylene (µg/m3)	**0.11(0.05-0.24)**	**0.001**	**0.04(0.01-0.08)**	**<0.001**	**19.89(5.19-76.24)**	**<0.001**
M-xylene (µg/m3)	**0.37(0.24-0.58)**	**<0.001**	**0.20(0.12-0.33)**	**<0.001**	**4.08(2.13-7.80)**	**<0.001**
TBTEX (µg/m3)	**0.89(0.82-0.97)**	**0.007**	**0.79(0.73-0.87)**	**<0.001**	**1.29(1.16-1.44)**	**<0.001**
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	**1.003(1.001-1.006)**	**0.001**	**1.004(1.001-1.007)**	**0.016**	**1.008(1.005-1.01)**	**<0.001**

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m^2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, life expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

Table 8. The estimated incidence rate ratios using negative binomial regression analyses for the effect of each 10 unit increase in air pollutants on total Leukemia incidence in adults across the districts of Tehran.

	Model 1		Model 2		Model 3	
	IRR (95% CI)	P-value	IRR (95% CI)	P-value	IRR (95% CI)	P-value
Pollutant						
Single-pollutant						
Annual PM_{10} (µg/m3)	**0.92(0.90-0.95)**	**<0.001**	**0.89(0.87-0.92)**	**<0.001**	1.01(0.97-1.05)	0.479
Annual SO_2 (ppb)	**0.96(0.94-0.99)**	**0.012**	0.97(0.95-1.01)	0.109	0.98(0.96-1.02)	0.390
Annual NO_2 (ppb)	**1.21(1.14-1.27)**	**<0.001**	**1.19(1.13-1.26)**	**<0.001**	**1.10(1.03-1.18)**	**0.003**
Annual NO (ppb)	**0.98(0.97-0.99)**	**0.028**	**0.97(0.95-0.98)**	**<0.001**	**1.03(1.01-1.05)**	**0.005**
Annual NO_X (ppb)	**1.02(1.00-1.03)**	**0.008**	**1.01(0.99-1.02)**	**0.338**	**1.03(1.01-1.04)**	**<0.001**
Benzene (µg/m3)	0.66(0.42-1.05)	0.078	**0.44(0.27-0.72)**	**0.001**	1.44(0.87-2.40)	0.158
Toluene (µg/m3)	**0.87(0.78-0.97)**	**0.017**	**0.79(0.71-0.89)**	**<0.001**	**1.19(1.05-1.37)**	**0.008**
Ethylbenzene (µg/m3)	**0.36(0.19-0.67)**	**0.001**	**0.20(0.11-0.38)**	**<0.001**	**2.28(1.05-4.94)**	**0.037**
P-xylene (µg/m3)	**0.39(0.19-0.80)**	**0.010**	**0.14(0.06-0.30)**	**<0.001**	1.98(0.86-4.55)	0.106
O-xylene (µg/m3)	0.33(0.19-0.55)	**<0.001**	**0.22(0.13-0.37)**	**<0.001**	**3.03(1.36-6.74)**	**0.006**
M-xylene (µg/m3)	0.59(0.44-0.78)	**<0.001**	**0.43(0.31-0.58)**	**<0.001**	1.39(0.95-2.02)	0.085
TBTEX (µg/m3)	**0.94(0.89-0.99)**	**0.042**	**0.88(0.83-0.94)**	**<0.001**	**1.08(1.01-1.15)**	**0.019**
Multi-pollutant						
profile 1 (low pollution)	Ref		Ref		Ref	
profile 2 (high pollution)	**1.002(1.00-1.003)**	**0.017**	0.99(0.99-1.001)	0.892	**1.003(1.001-1.005)**	**<0.001**

Model 1: Adjusted for age and gender.
Model 2: Adjusted for age, gender and urban green space per capita (m^2 per 1000 people).
Model 3: Adjusted for age, gender, socioeconomic status, life expectancy.
The incidence rate ratio of Leukemia is estimated for each 10 unit increase in pollutants.

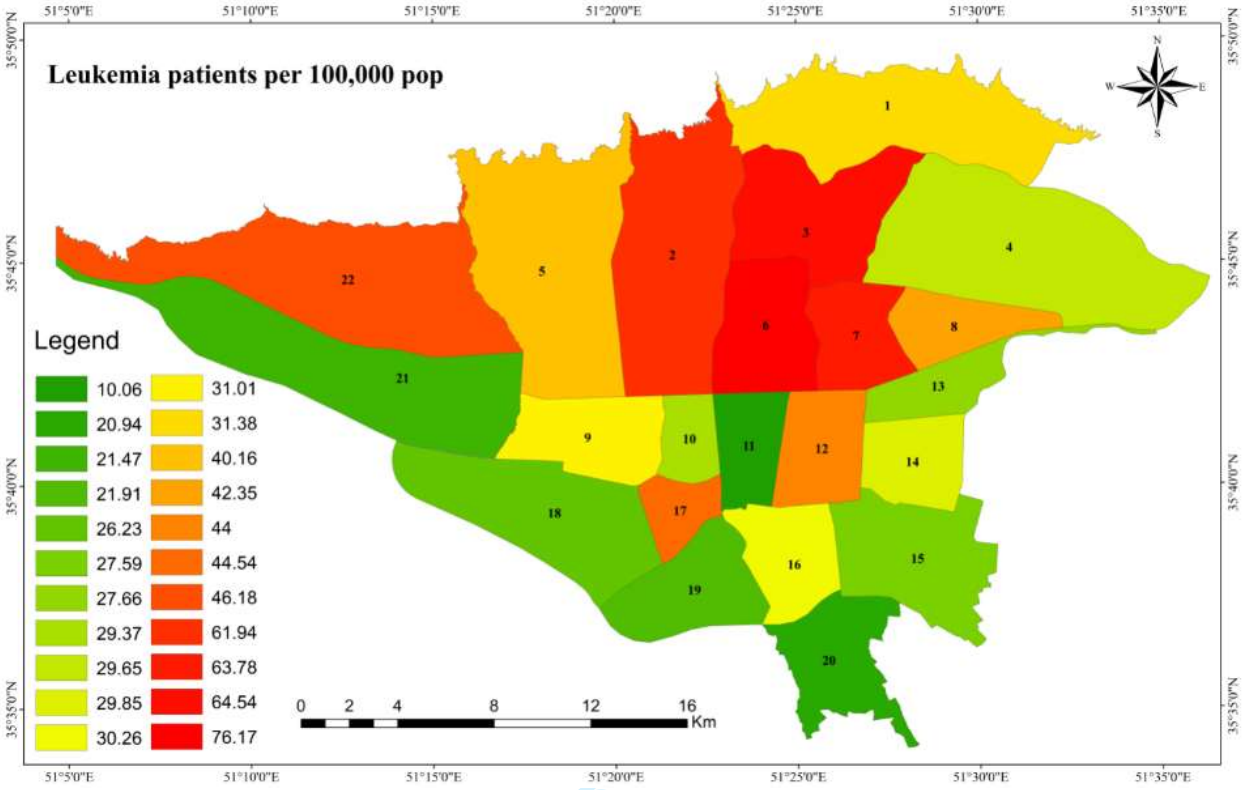

Fig. 1. Spatial distribution of Leukemia cancer patients (number of cases in 100,000) in different areas of Tehran in 2014-2016 (The labels on the map are district numbers).

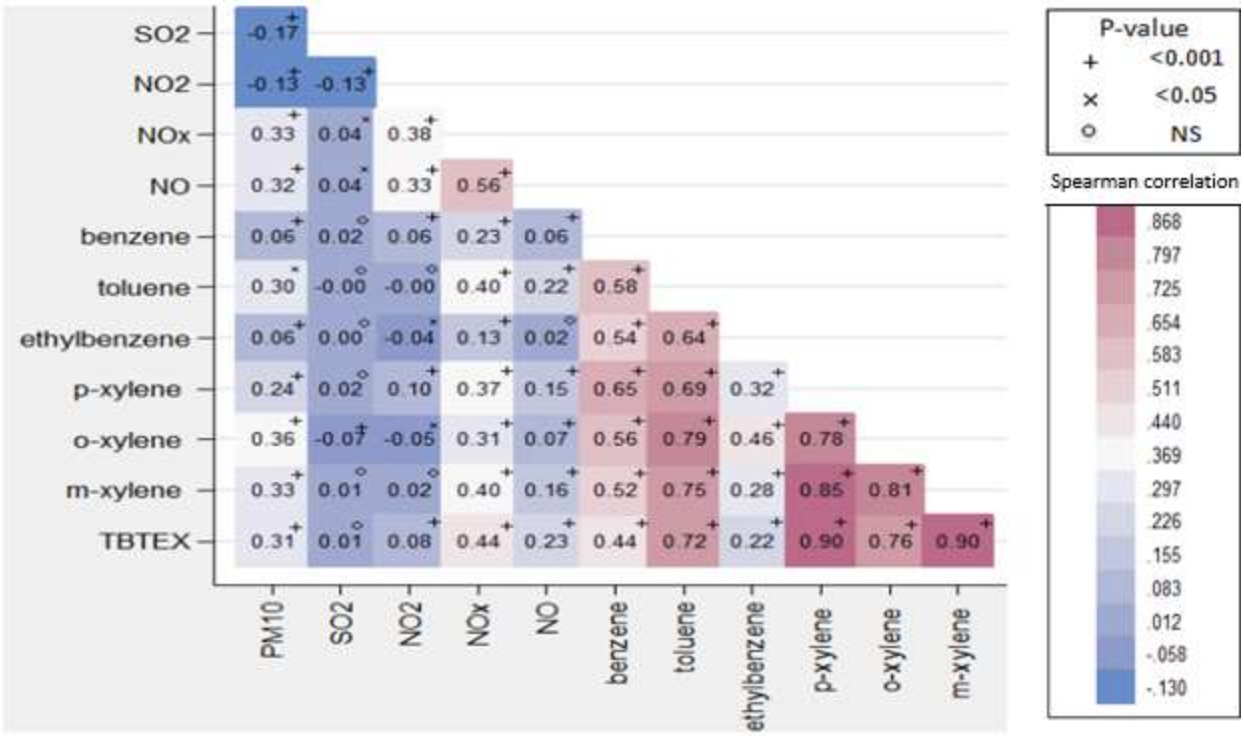

Fig. 2. Spearman correlation matrix of air pollutants in the 22 districts of Tehran, Iran, in 2014-2016.

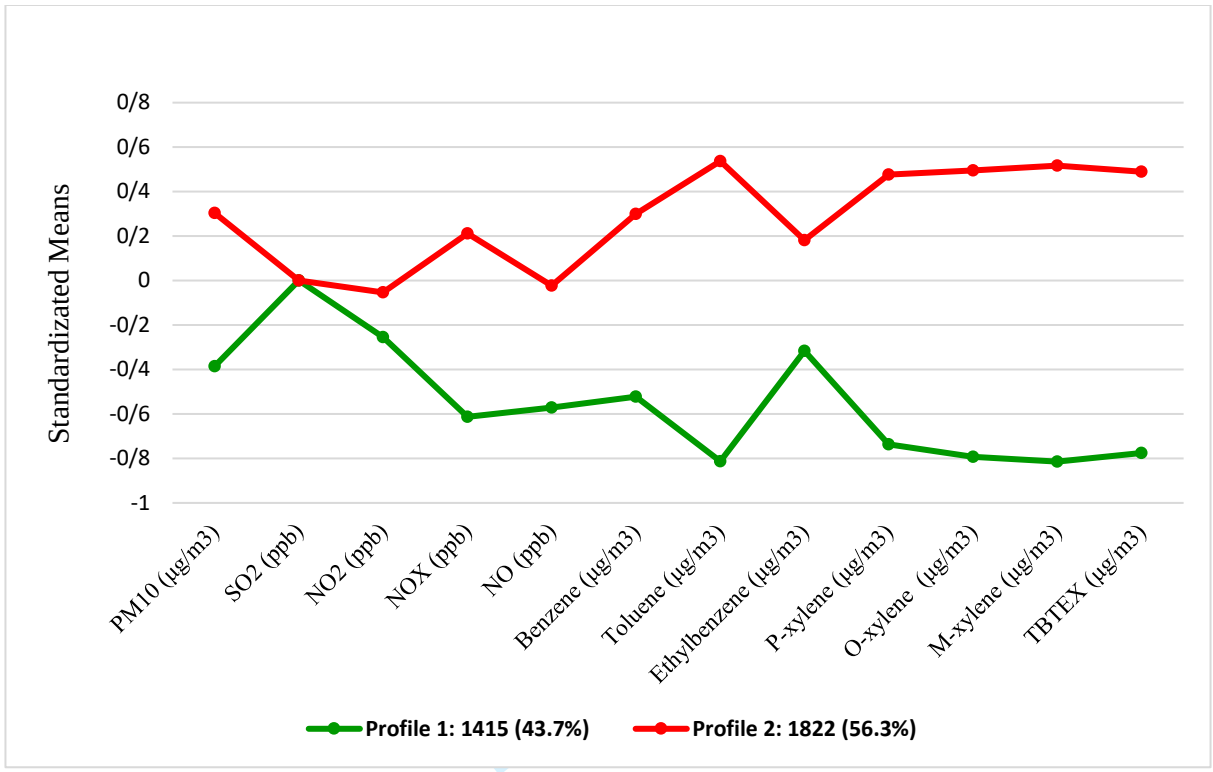

Fig. 3. Standard mean values of pollutants in the two latent profiles in different areas of Tehran, Iran, in 2014-2016.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4,5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4,5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4,5*
Bias	9	Describe any efforts to address potential sources of bias	5
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5,6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5,6
		(*b*) Describe any methods used to examine subgroups and interactions	5,6
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	NA
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	6
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*NA*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	*NA*
		Cross-sectional study—Report numbers of outcome events or summary measures	*6, 16*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	22
		(*b*) Report category boundaries when continuous variables were categorized	16
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	22,23,24
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	23,24
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8,9
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	11

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.