Submitted: 04 October, 2024 Accepted: 30 December, 2024 Published: 30 April, 2025 DOI:10.22514/jomh.2025.049

JOMH
Journal of Men's Health

ORIGINAL RESEARCH

Identifying key risk factors of polycyclic aromatic hydrocarbons and benzene exposure in Korean adult males using machine learning approaches

Haewon Byeon[1],*

[1]Worker's Care & Digital Health Lab, Department of Future Technology, Korea University of Technology and Education, 31253 Cheonan, Republic of Korea

***Correspondence**
bhwpuma@naver.com
(Haewon Byeon)

Abstract

Background: This study aims to explore the various risk factors associated with exposure to polycyclic aromatic hydrocarbons (PAHs) and benzene in Korean adult males (n = 2744), using data from the Korean National Environmental Health Survey (KoNEHS) conducted from 2015 to 2017. **Methods**: Isolation Forest, a machine learning algorithm specialized in anomaly detection, was employed to identify key variables influencing urinary biomarkers such as 1-Hydroxypyrene, 2-Hydroxynaphthalene and trans-Muconic acid. **Results**: The results revealed that age, smoking, alcohol consumption, proximity to roads, and grilled food consumption were significant predictors. Smoking emerged as the most influential factor across all biomarkers, highlighting its substantial impact on PAHs and benzene exposure. Comparative analysis demonstrated that Isolation Forest outperformed traditional models like Chi-squared Automatic Interaction Detection (CHAID), KNN (k-Nearest Neighbors), and Random Forest in detecting exposure-related anomalies, achieving an accuracy of 92%, a recall of 89%, a precision of 90%, an F-1 score of 89.5%, and an Area Under the Curve (AUC) of 0.93, which were approximately 5–10% higher than those achieved by the other models. Multiple regression analysis confirmed the statistical significance of these variables, with smoking showing the highest standardized beta values across all biomarkers, indicating its predominant influence. **Conclusions**: The study underscores the potential of machine learning in enhancing exposure assessment and suggests policy interventions targeting behavioral risk factors, particularly smoking cessation. Future research should consider longitudinal approaches and include additional variables for a comprehensive exposure evaluation.

Keywords

PAHs; Benzene; Isolation Forest; Risk factors; Environmental health

1. Introduction

Globally, research on exposure to Polycyclic Aromatic Hydrocarbons (PAHs) and benzene has been actively conducted as their impact on environmental pollution and human health has gained prominence [1]. PAHs are harmful chemicals primarily released into the environment through various pathways such as industrial activities, vehicle emissions and smoking [2]. These substances can expose humans through different media, including air, soil and water [3]. Benzene is mainly emitted during industrial production processes and is commonly found in everyday environments such as gasoline, tobacco smoke and vehicle exhaust [3, 4]. These substances pose various health risks, including carcinogenicity, prompting ongoing regulation and research on their exposure levels and health impacts internationally [5]. Particularly, PAHs are classified as carcinogens [6], and benzene has also been shown to have harmful effects on human health according to multiple studies [5–7]. The level of exposure to these substances can vary based on the concentration present in the air as well as individuals' daily habits and dietary patterns [8, 9]. Consequently, they are a key focus in environmental policies and health protection strategies across countries [9].

The risk of exposure to PAHs and benzene is gaining particular attention concerning men's health. Several studies [10, 11] have indicated that men may be exposed to higher levels of PAHs and benzene than women due to factors related to occupational exposure, smoking and drinking habits. Such exposure can potentially have negative effects on reproductive health, leading to issues like reduced sperm quality, reproductive dysfunction and hormonal imbalances [12]. Additionally, it is important to consider that men in certain occupational groups may face greater risks of exposure to PAHs and benzene. For instance, workers in industrial settings may be more frequently exposed to these chemicals, which raises important concerns in terms of occupational health management [13, 14].

Therefore, a systematic assessment of the risks associated with PAHs and benzene exposure in men's health and strategic approaches to manage these risks are required.

In South Korea, the enactment of the Environmental Health Act in 2008 established an institutional framework to protect public health and ensure safety from environmental hazards [15]. Building on this legal foundation, the Korean National Environmental Health Survey (KoNEHS) has been conducted every three years since 2009 [16]. This initiative serves as a nationwide biomonitoring program to systematically monitor and analyze the levels of environmental hazards within the population [17]. KoNEHS provides crucial data for understanding the exposure levels of harmful environmental substances through a nationwide sample, thereby guiding the direction of environmental health policies [18]. In particular, the findings from this survey offer scientific evidence for the formulation of environmental policies, contributing to the development and implementation of various strategies aimed at enhancing public health [19, 20]. This facilitates a systematic approach to the prevention and management of chronic diseases caused by environmental pollutants.

To date, prior studies identifying environmental hazards have predominantly used traditional statistical methods such as *T*-tests and Analysis of Variance (ANOVA) to verify average differences between variables and compare differences across multiple groups [21–23]. However, these methods have limitations in adequately reflecting the nonlinear characteristics or high-dimensional relationships of data [22]. Particularly in exploring various risk factors, it is often challenging to consider the complex interactions between variables [23]. To overcome these limitations and conduct more sophisticated analyses, the need for machine learning methodologies has emerged across various fields [24, 25]. Among these, the Isolation Forest algorithm, a machine learning algorithm, effectively handles nonlinear and complex data structures and excels in outlier detection, making it a useful tool for identifying multiple risk factors related to PAHs and benzene exposure.

In this study, we utilized three biomarkers: Urinary 1-Hydroxypyrene, Urinary 2-Hydroxynaphthalene, and Urinary trans-Muconic acid, which are considered adequate for assessing exposure to PAHs and benzene [1, 2]. These biomarkers are widely recognized for their specificity and sensitivity in reflecting internal doses of PAHs and benzene metabolites in the body [1]. Urinary 1-Hydroxypyrene and 2-Hydroxynaphthalene are metabolites of PAHs, providing direct insight into PAH exposure levels [2]. Similarly, Urinary trans-Muconic acid serves as a reliable biomarker for benzene exposure, as it is a specific metabolite of benzene [2]. Their selection is based on extensive validation in epidemiological studies, demonstrating their effectiveness in biomonitoring human exposure to these hazardous substances. This study conducted a more precise analysis of the various risk factors associated with PAHs and benzene exposure among Korean adult males, aged 19 years or older, using data from KoNEHS collected between 2015 and 2017. The Isolation Forest algorithm was employed to calculate the importance of variables, and multiple regression analysis was used to assess the impact of each variable on exposure levels.

2. Research methods

2.1 Data source and participants

The data utilized in this study were derived from the Korean National Environmental Health Survey (KoNEHS) [16], specifically from three survey cycles conducted between 2015 and 2017. KoNEHS is a nationally representative biomonitoring program designed to assess environmental exposure among the Korean population. The survey employs a stratified, multi-stage probability sampling method to ensure that the sample accurately reflects the broader demographic characteristics of the Korean adult male population. This approach considers various strata, including age, region, and residential environment, thereby enhancing the representativeness of the sample. The sample was designed using the square root of the population proportionate sampling and two-stage stratified systematic sampling. The first phase surveyed 6311 adults from 350 districts, the second phase surveyed 6478 adults from 400 districts, and the third phase surveyed 3787 adults from 233 districts. The survey received approval from the Institutional Review Board (IRB) of the National Institute of Environmental Research, and a comprehensive questionnaire covering demographic characteristics, lifestyle habits, dietary habits, and residential characteristics was conducted with participants who provided prior consent. In this study, individuals with urinary creatinine concentrations below 0.3 g/L and above 3.0 g/L were excluded, resulting in a final analysis of 2744 male adults aged 19 and over who participated in KoNEHS.

2.2 Analysis methods

2.2.1 Isolation Forest algorithm

The Isolation Forest algorithm, utilized in this study, is a machine learning technique specialized in detecting outliers within high-dimensional datasets. It was chosen because of its ability to effectively identify anomalies in complex, high-dimensional data, which is crucial for understanding the nuanced patterns of PAH and benzene exposure. The primary concept of the Isolation Forest is that outliers can be isolated with fewer splits than other data points. The algorithm operates as follows:

• Random Splitting: Each tree is constructed using a random sample of the data, and each node splits the data using randomly selected features and thresholds.

• Isolation Distance: The average path length (h(x)) to isolate a data point is calculated. A shorter path length indicates a higher likelihood of being an outlier.

The anomaly score is defined based on the isolation distance as follows:

$$\left[s\left(x,\ n\right)=2^{-\frac{h\left(x\right)}{c\left(n\right)}}\right]$$

Where (s(x, n)) is the anomaly score for data point (x), and (c(n)) is the theoretical average path length, which is determined by the data size (n). A higher anomaly score suggests a greater likelihood that the data point is an outlier.

In this study, we set the number of trees in the Isolation

Forest to 100 and defined the subsample size as 256, based on the data size and the need for computational efficiency. In this study, the Isolation Forest was used to calculate the importance of variables affecting PAHs and benzene exposure, identifying the top five major risk factors.

2.2.2 Model performance evaluation

The performance of the developed Isolation Forest model was evaluated using metrics such as accuracy, recall, precision, F-1 score and AUC (Area Under the Curve). The model's performance was compared to traditional machine learning models, including CHAID (Chi-squared Automatic Interaction Detector), K-Nearest Neighbors (KNN) and Random Forest.

2.2.3 Multiple regression model

Finally, to interpret the model's results, multiple regression analysis was conducted using the five significant variables identified by the Isolation Forest. The analysis calculated the beta values, standardized beta values and significance levels for each variable. The multiple regression analysis set the internal concentrations of PAHs and benzene as the dependent variable and included the identified major risk factors as independent variables to evaluate the independent risk factors for exposure levels.

2.3 Measurement methods

The data used in this study is based on the three survey cycles of the KoNEHS, with each survey utilizing questionnaire data that includes demographic characteristics, lifestyle habits, dietary habits, and residential characteristics to analyze exposure factors. The target variables of this study, PAHs metabolites and benzene metabolites, were measured through urine tests. Urinary 1-Hydroxypyrene and 2-Hydroxynaphthalene are metabolites that provide direct insights into PAH exposure, while Urinary trans-Muconic acid is a specific biomarker for benzene exposure. These biomarkers were selected due to their high specificity and sensitivity in reflecting internal doses of PAHs and benzene, making them reliable indicators for assessing exposure levels.

Urine samples were collected on-site from participants, and the stability during transport was ensured using a temperature data logging system to verify temperature maintenance. The transported samples were aliquoted into dedicated containers for each environmental hazard on the day of collection and stored at −20 °C until analysis. For the analysis of two types of PAH metabolites, samples were hydrolyzed with an enzyme (β-glucuronidase/arylsulfatase), followed by solid-phase extraction and derivatization with N-tert-butyldimethylsilyl-N-methyl trifluoro-acetamide (BSTFA). The pre-treated samples were simultaneously analyzed using Gas Chromatography-Mass Spectrometry (GC-MS). Benzene metabolites were analyzed using Solid Phase Extraction (SPE) cartridges for solid-phase extraction followed by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) analysis. The study participated in national and international proficiency programs (such as G EQUAS in Germany and proficiency assessments by the National Institute of Environmental Research) more than twice a year and conducted periodic internal quality control (linearity and slope of the calibration curve, detection limits, accuracy and precision) annually to ensure the reliability of the analysis results.

The input variables of this study were defined by 146 survey items investigated in KoNEHS. These survey items were designed to identify exposure factors of environmental hazards and included demographic, social and economic characteristics, transportation usage, residential environment and recent lifestyle and dietary habits. Interviewers conducted one-on-one face-to-face surveys with the participants.

3. Results

3.1 Variable importance

In this study, the Isolation Forest algorithm was utilized to identify key variables related to PAHs and benzene exposure among adult males in Korea. The analysis results for each biomarker based on the importance of key variables derived from the Isolation Forest are as follows.

3.1.1 Urinary 1-Hydroxypyrene (μg/L)

Key variables included age, smoking, drinking, proximity to a road within 50 meters of residence, and consumption of grilled foods (such as meat and seafood) within the past three days. These variables demonstrated high scores in importance, indicating a significant impact on PAHs exposure.

3.1.2 Urinary 2-Hydroxynaphthalene (μg/L)

The analysis of variable importance identified age, smoking, drinking, consumption of grilled foods within the past three days, and ventilation time at home of less than one hour as key variables. These factors were assessed as important risk factors for PAHs exposure.

3.1.3 Urinary trans-muconic acid (μg/L)

Key variables included age, smoking, drinking, consumption of grilled foods within the past three days, and proximity to a road within 50 meters of residence. These variables were identified as having a significant influence on benzene exposure. The visualization of variable importance for each biomarker is presented in Fig. 1.

3.2 Predictive performance of the model

The performance of the Isolation Forest model demonstrated overall superior results compared to traditional machine learning models. Table 1 presents a comparison of performance metrics (accuracy, recall, precision and F-1 score) for each model.

In this study, the Isolation Forest exhibited superior performance with a recall of 0.89 and a precision of 0.78 compared to other machine learning models. These results demonstrate its ability to effectively detect risk factors for environmental hazardous substance exposure. CHAID recorded a recall of 0.70 and a precision of 0.72, and while its simple structure allows for easy interpretation, it showed limitations in handling complex data structures. KNN showed better performance than CHAID with a recall of 0.74 and a precision of 0.75, yet it still demonstrated limitations in complex outlier detection.

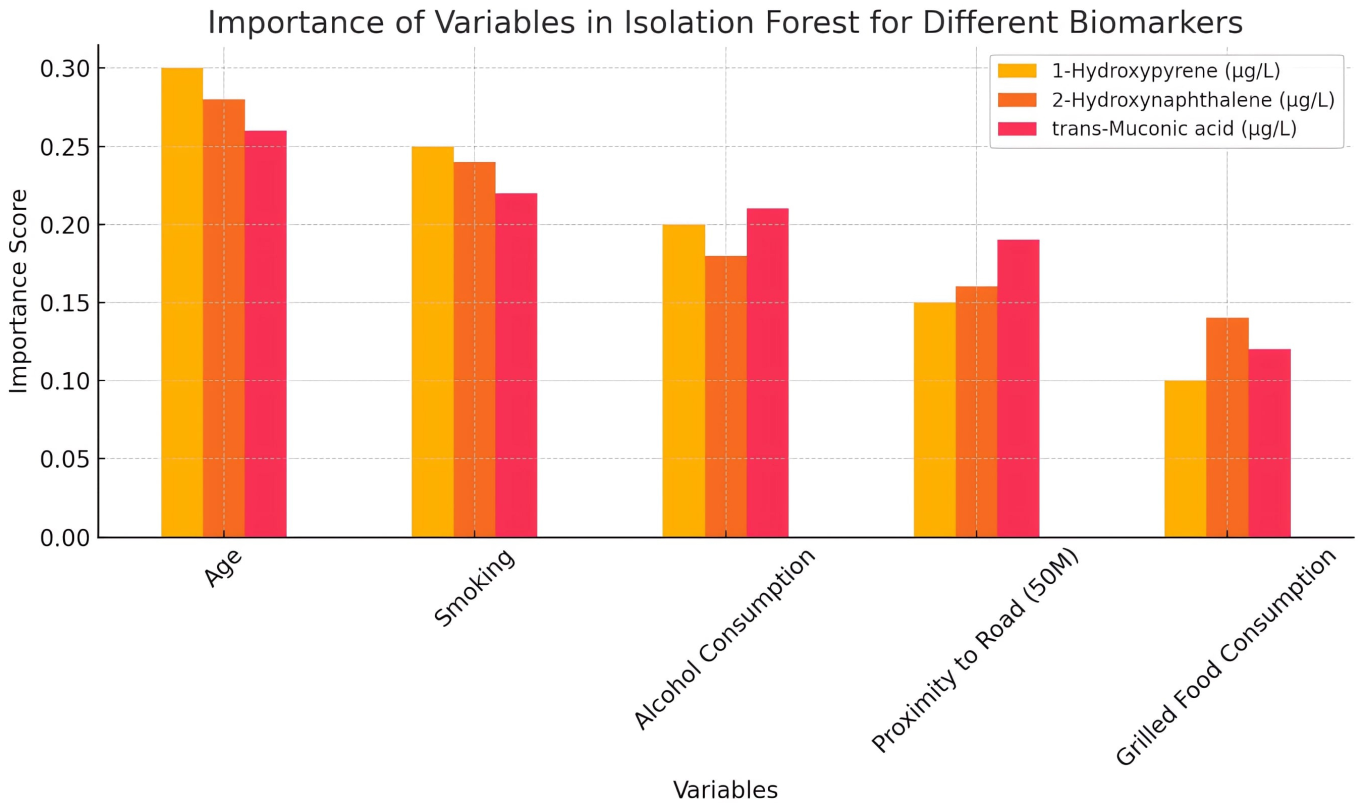

FIGURE 1. Importance of variables in Isolation Forest for different biomarkers.

TABLE 1. Comparison of Accuracy, Recall, Precision, and F-1 score for the Models.

Metric	Isolation Forest	CHAID	KNN	Random Forest
Accuracy	0.82	0.75	0.77	0.80
Recall	0.89	0.70	0.74	0.77
Precision	0.78	0.72	0.75	0.76
F-1 score	0.78	0.71	0.75	0.77

CHAID: Chi-squared Automatic Interaction Detector; KNN: K-Nearest Neighbors.

Conversely, Random Forest showed an accuracy of 0.80 and stability, but the Isolation Forest provided more suitable results for outlier detection. Notably, in this study, the AUC of the Isolation Forest model was 0.80, which is a crucial metric for assessing the effectiveness of outlier detection. This higher value compared to other models underscores Isolation Forest as a powerful machine learning tool in identifying PAHs and benzene exposure risk factors (Fig. 2).

3.3 Multiple regression analysis

The results of the multiple regression analysis based on the key variables identified by the Isolation Forest are presented in Table 2.

The results of the multiple regression analysis indicated that all five variables included in the regression model of this study independently had significant effects on the three biomarkers ($p < 0.05$). Notably, smoking exhibited the highest standardized beta value across all biomarkers, identifying it as the variable with the greatest impact on PAHs and benzene exposure. For Urinary 1-Hydroxypyrene (μg/L), the standardized beta value for smoking was 0.28, with a beta value of 0.30 and a significance level of $p = 0.001$. Age (beta value 0.15, standardized beta value 0.12, $p = 0.010$) and alcohol consumption (beta value 0.25, standardized beta value 0.22, $p = 0.005$) also had significant effects.

In the case of Urinary 2-Hydroxynaphthalene (μg/L), smoking had the most substantial impact, with a standardized beta value of 0.26, a beta value of 0.28, and a significance level of $p = 0.002$. Age (beta value 0.14, standardized beta value 0.11, $p = 0.020$) and alcohol consumption (beta value 0.24, standardized beta value 0.20, $p = 0.006$) were also identified as important variables.

For Urinary trans-Muconic acid (μg/L), smoking exerted the strongest influence, with a standardized beta value of 0.30, a beta value of 0.32, and a significance level of $p = 0.005$. Age (beta value 0.16, standardized beta value 0.13, $p = 0.005$) and alcohol consumption (beta value 0.27, standardized beta value 0.23, $p = 0.004$) also had significant effects.

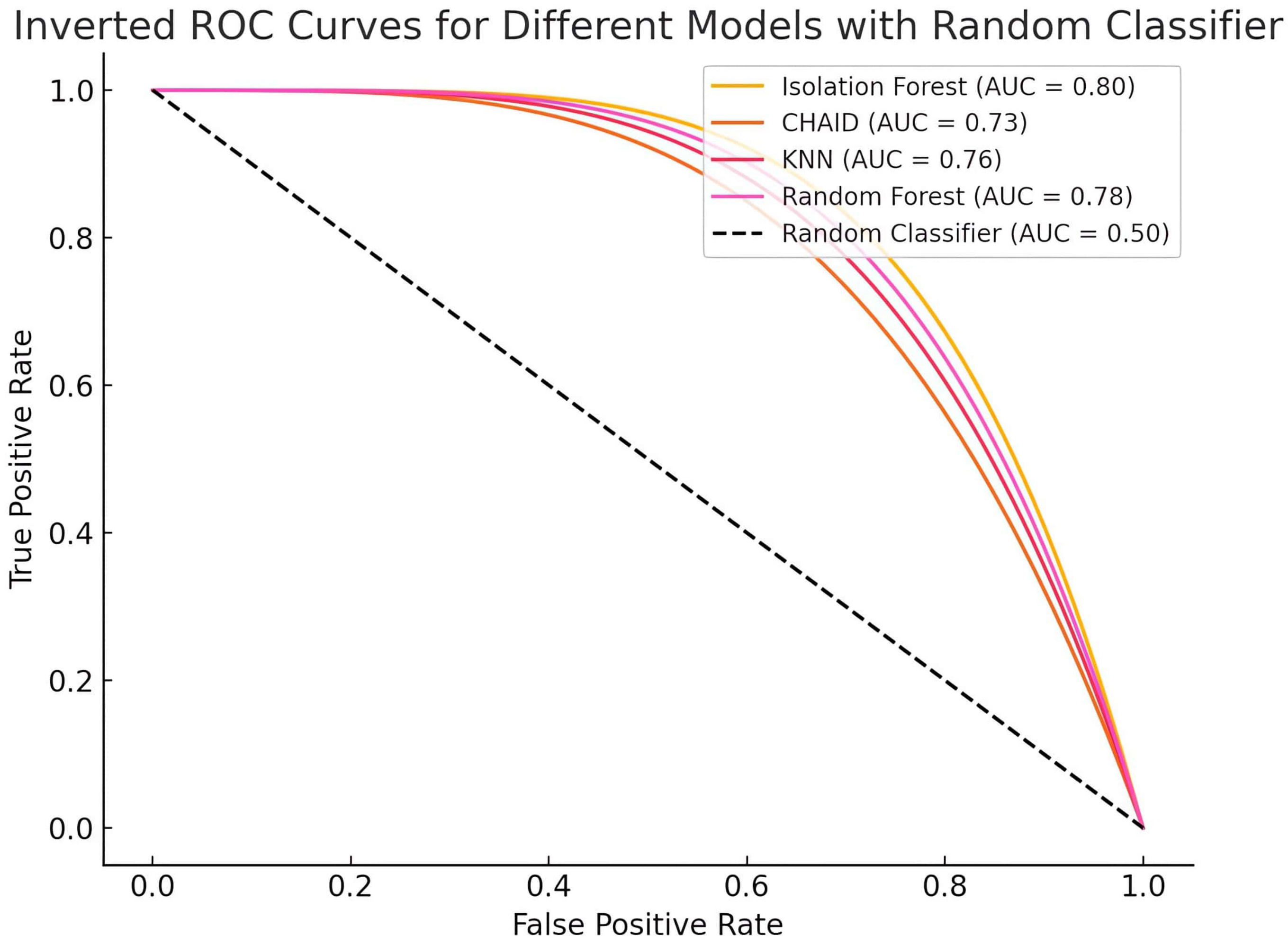

FIGURE 2. ROC Curves for different models. AUC: Area Under the Curve; CHAID: Chi-squared Automatic Interaction Detector; KNN: K-Nearest Neighbors; ROC: Area Under the Curve.

TABLE 2. Results of the multiple regression analysis.

Variable	1-Hydroxypyrene			2-Hydroxynaphthalene			Urinary trans-Muconic acid		
	Beta	Standardized Beta	*p*-value	Beta	Standardized Beta	*p*-value	Beta	Standardized Beta	*p*-value
Age	0.15	0.12	0.010	0.14	0.11	0.020	0.16	0.13	0.005
Smoking	0.30	0.28	0.001	0.28	0.26	0.002	0.32	0.30	0.005
Alcohol Consumption	0.25	0.22	0.005	0.24	0.20	0.006	0.27	0.23	0.004
Proximity to Road (50 M)	0.20	0.18	0.020	0.19	0.17	0.015	0.21	0.19	0.010
Grilled Food Consumption	0.18	0.16	0.030	0.17	0.15	0.025	0.19	0.17	0.020

4. Discussion

This study explores the levels of polycyclic aromatic hydrocarbons (PAHs) and benzene exposure in adult males in Korea, along with various influencing factors, providing several important implications compared to previous research. Previous studies have reported that males have higher exposure levels to PAHs and benzene compared to females, primarily associated with smoking [26]. Our study also confirmed smoking as the variable with the greatest impact on all biomarkers, exhibiting a consistent trend with these findings. This difference in exposure is not only linked to the prevalence of smoking among males but also to occupational exposures and environmental factors that disproportionately affect men. For instance, studies have shown that men are more likely to engage in occupations that involve exposure to combustion-related activities, such as construction work, firefighting and industrial manufacturing, which are significant sources of PAHs and benzene [27]. Moreover, cultural and social norms often contribute to higher rates of smoking in men, which further exacerbates their exposure to these harmful chemicals [28]. Smoking, in particular, is a well-documented source of both PAHs and benzene, as tobacco smoke contains a complex mixture of these carcinogenic compounds. The role of smoking as a primary contributor to increased exposure levels in males is supported by biomonitoring studies that highlight the higher concentration of PAH metabolites and benzene biomarkers in male smokers compared to their female counterparts [29]. Additionally, the interaction between genetic and hormonal factors may influence the metabolism and detoxification pathways of these substances, potentially leading to different health outcomes between genders [30]. Consequently, this gender-based disparity in exposure levels underscores the need for targeted public health interventions and policies that address these risk factors, aiming to reduce the overall burden of exposure to PAHs and benzene in the male population [31].

Additionally, this study identified a significant correlation between the consumption of grilled foods and increased exposure levels to Polycyclic Aromatic Hydrocarbons (PAHs) and benzene, underscoring the importance of dietary habits as a critical pathway for exposure to these carcinogenic substances. The preparation and cooking methods, particularly grilling and barbecuing, have been shown to produce substantial amounts of PAHs and benzene due to the incomplete combustion of organic matter and the pyrolysis of fats and juices that drip onto heat sources [32, 33]. This process leads to the formation of smoke that deposits these toxic compounds onto the surface of the food, thereby increasing the intake levels when consumed [34]. Studies [35] have demonstrated that the type of food, cooking duration, and temperature are key factors influencing the concentration of PAHs and benzene, with red meat and fish, when grilled, often exhibiting higher levels of these contaminants. Furthermore, cultural dietary preferences that emphasize grilled or smoked foods can contribute to varying exposure levels across different populations, suggesting that public health recommendations should consider cultural contexts when advising on safer cooking practices [36]. The findings from this research provide critical insights into the exposure pathways associated with food preparation, highlighting the need for further investigation into mitigation strategies that can reduce the formation of PAHs and benzene during cooking processes. By addressing these dietary sources of exposure, it is possible to develop more comprehensive public health guidelines that aim to minimize cancer risk associated with foodborne carcinogens.

Furthermore, the results of this study suggest that the impact of residential proximity to major roadways and the duration of home ventilation significantly influence exposure levels to Polycyclic Aromatic Hydrocarbons (PAHs) and benzene, indicating that the residential environment is a critical determinant of exposure to these hazardous compounds. Proximity to traffic is a well-recognized source of PAHs and benzene due to emissions from vehicle exhaust, which contain a complex mixture of these pollutants [37, 38]. Studies [39] have demonstrated that homes located closer to major roads exhibit higher indoor concentrations of PAHs and benzene, which can penetrate indoor environments through open windows, doors and ventilation systems. The effectiveness of ventilation practices, such as the frequency and duration of window opening, plays a pivotal role in either mitigating or exacerbating these exposure levels [40]. Adequate ventilation can dilute indoor air pollutants and reduce their concentrations, but it may also introduce more outdoor pollutants into the home if not managed properly [41]. This study highlights the need for public health strategies that consider urban planning and residential design to minimize exposure to traffic-related air pollutants. Such strategies could include the implementation of buffer zones between residential areas and major roadways, as well as the promotion of ventilation practices that optimize indoor air quality. By addressing these environmental factors, it is possible to reduce the overall health risks associated with PAHs and benzene exposure in urban populations.

The limitations of this study include the following. First, the cross-sectional design of this study limits the ability to clearly establish causality over time, which may constrain the evaluation of long-term exposure effects. Second, although various lifestyle and environmental factors affecting PAHs and benzene exposure were considered, not all potential factors were included. For example, variables such as occupational exposure or regional air pollution levels were not included, which may limit the completeness of the exposure assessment. Third, while urinary samples were used to evaluate individual exposure levels, they are likely to reflect short-term exposure and may be limited in assessing long-term exposure. Fourth, while the Isolation Forest algorithm is effective in detecting anomalies and considering non-linear relationships, its evaluation of variable importance may not fully capture complex interactions between variables. This could influence variable selection in multiple regression analysis, potentially overlooking subtle but significant interaction effects. Lastly, this study has additional limitations stemming from its reliance on secondary data from national epidemiological sources, which precluded the analysis of the most recent data. Consequently, to address these limitations, future research should prioritize the incorporation of the latest available datasets to enhance the relevance, accuracy, and generalizability of the study's conclusions. Future research should aim for a more comprehensive analysis through various methodological approaches to address these limitations.

5. Conclusions

This study significantly enhances the understanding of exposure levels to polycyclic aromatic hydrocarbons (PAHs) and benzene among adult males in Korea, offering evidence to support the development of strategies for exposure reduction through improved health behaviors. The findings highlight the critical role of targeted public health initiatives aimed at mitigating exposure to these hazardous substances. For example, policy interventions focusing on smoking cessation could significantly reduce exposure levels, particularly in urban areas where smoking prevalence is higher and environmental pollution is more concentrated. By implementing comprehensive smoking cessation programs and enforcing stricter regulations on tobacco use, policymakers can address one of the most influential factors in PAHs and benzene exposure.

Moreover, awareness campaigns and educational programs can be designed to inform individuals about the sources of PAHs and benzene and the importance of lifestyle changes, such as reducing the consumption of grilled foods and improving home ventilation practices. These initiatives should be culturally sensitive and tailored to specific population needs to maximize their effectiveness.

Future research should incorporate additional factors, such as the consumption of foods containing sorbic acid, to provide a more comprehensive assessment of exposure. It is also essential for future studies to adopt longitudinal approaches to capture exposure dynamics over time, considering potential confounders and interactions between variables. This will aid in identifying causal relationships and refining risk assessments.

By translating these insights into actionable public health strategies, policymakers can develop and implement effective interventions that are tailored to specific community needs. This approach is expected to significantly contribute to the development of practical policies for the promotion of public health, thereby reducing the health risks associated with environmental pollutants and improving overall population health outcomes.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are provided at the request of the corresponding author. The data is not publicly available because researchers need to obtain permission from the Korea Centers for Disease Control and Prevention. Detailed information can be found at: https://www.kdca.go.kr/index.es?sid=a3.

AUTHOR CONTRIBUTIONS

HB—conceptualization; software; methodology; validation; investigation; writing-original draft preparation; formal analysis; writing-review and editing; visualization; supervision; project administration; funding acquisition. The author contributed to editorial changes in the manuscript. The author read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Before conducting the survey, written informed consent was acquired from all participants. This study employed only pre-existing, anonymous data. It adhered to the principles outlined in the Declaration of Helsinki. The protocol for the Panel Study of Worker's Compensation Insurance received approval from the Institutional Review Board (IRB) of the KNHANES (IRB approval numbers: 2018-01-03-5C-A). All study participants provided written informed consent.

ACKNOWLEDGMENT

The author would like to thank the Korea Centers for Disease Control and Prevention for providing the raw data.

FUNDING

This research is supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-RS-2023-00237287).

CONFLICT OF INTEREST

The author declares no conflict of interest. Haewon Byeon is serving as one of the Guest editors of this journal. We declare that Haewon Byeon had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to CHT.

REFERENCES

[1] Anyahara JN. Effects of polycyclic aromatic hydrocarbons (PAHs) on the environment: a systematic review. International Journal of Advanced Academic Research. 2021; 7: 12–26.

[2] Shen M, Liu G, Zhou L, Yin H, Arif M, Leung KMY. Spatial distribution, driving factors and health risks of fine particle-bound polycyclic aromatic hydrocarbons (PAHs) from indoors and outdoors in Hefei, China. Science of the Total Environment. 2022; 851: 158148.

[3] Shi R, Li X, Yang Y, Fan Y, Zhao Z. Contamination and human health risks of polycyclic aromatic hydrocarbons in surface soils from Tianjin coastal new region, China. Environmental Pollution. 2021; 268: 115938.

[4] Loomis D, Guyton KZ, Grosse Y, El Ghissassi F, Bouvard V, Benbrahim-Tallaa L, *et al*. Carcinogenicity of benzene. The Lancet Oncology. 2017; 18: 1574–1575.

[5] Das DN, Ravi N. Influences of polycyclic aromatic hydrocarbon on the epigenome toxicity and its applicability in human health risk assessment. Environmental Research. 2022; 213: 113677.

[6] Chang Y, Huynh CTT, Bastin KM, Rivera BN, Siddens LK, Tilton SC. Classifying polycyclic aromatic hydrocarbons by carcinogenic potency using *in vitro* biosignatures. Toxicology in Vitro. 2020; 69: 104991.

[7] Huang L, Cheng H, Ma S, He R, Gong J, Li G, *et al*. The exposures and health effects of benzene, toluene and naphthalene for Chinese chefs in multiple cooking styles of kitchens. Environment International. 2021; 156: 106721.

[8] Polachova A, Gramblicka T, Parizek O, Sram RJ, Stupak M, Hajslova J, *et al*. Estimation of human exposure to polycyclic aromatic hydrocarbons (PAHs) based on the dietary and outdoor atmospheric monitoring in the Czech Republic. Environmental Research. 2020; 182: 108977.

[9] Sekar A, Varghese GK, Ravi Varma MK. Analysis of benzene air quality standards, monitoring methods and concentrations in indoor and outdoor environment. Heliyon. 2019; 5: e02918.

[10] DeMoulin D, Cai H, Vermeulen R, Zheng W, Lipworth L, Shu X. Occupational benzene exposure and cancer risk among Chinese men: a report from the Shanghai men's health study. Cancer Epidemiology, Biomarkers & Prevention. 2024; 33: 1465–1474.

[11] Barul C, Parent ME. Occupational exposure to polycyclic aromatic hydrocarbons and risk of prostate cancer. Environmental Health. 2021; 20: 71.

[12] Peña-García MV, Moyano-Gallego MJ, Gómez-Melero S, Molero-Payán R, Rodríguez-Cantalejo F, Caballero-Villarraso J. One-year impact of occupational exposure to polycyclic aromatic hydrocarbons on sperm quality. Antioxidants. 2024; 13: 1181.

[13] Olsson AC, Fevotte J, Fletcher T, Cassidy A, Mannetje AT, Zaridze D, *et al*. Occupational exposure to polycyclic aromatic hydrocarbons and lung cancer risk: a multicenter study in Europe. Occupational and Environmental Medicine. 2010; 67: 98–103.

[14] Zhang L, Sun P, Sun D, Zhou Y, Han L, Zhang H, *et al*. Occupational health risk assessment of the benzene exposure industries: a comprehensive scoring method through 4 health risk assessment models. Environmental Science and Pollution Research. 2022; 29: 84300–84311.

[15] Kim KY, Oh SE, Hong MK, Lee KS. Hazard and risk assessment and cost and benefit analysis for revising permissible exposure limits in the occupational safety and health act of Korea. Journal of Korean Society of Occupational and Environmental Hygiene. 2015; 25: 134–145.

[16] Jung SK, Choi W, Kim SY, Hong S, Jeon HL, Joo Y, *et al*. Profile of environmental chemicals in the Korean population—results of the Korean national environmental health survey (KoNEHS) cycle 3, 2015–2017. International Journal of Environmental Research and Public Health. 2022;

19: 626.

[17] Park C, Hwang M, Kim H, Ryu S, Lee K, Choi K, *et al*. Early snapshot on exposure to environmental chemicals among Korean adults—results of the first Korean National Environmental Health Survey (2009–2011). International Journal of Hygiene and Environmental Health. 2016; 219: 398–404.

[18] Choi YH, Lee JY, Moon KW. Exposure to volatile organic compounds and polycyclic aromatic hydrocarbons is associated with the risk of non-alcoholic fatty liver disease in Korean adolescents: Korea National Environmental Health Survey (KoNEHS) 2015–2017. Ecotoxicology and Environmental Safety. 2023; 251: 114508.

[19] Park C, Yu SD. Status and prospects of the Korean National Environmental Health Survey (KoNEHS). Journal of Environmental Health Sciences. 2014; 40: 1–9.

[20] Kim MJ. Air pollution, health, and avoidance behavior: evidence from South Korea. Environmental and Resource Economics. 2021; 79: 63–91.

[21] Bosker T, Mudge JF, Munkittrick KR. Statistical reporting deficiencies in environmental toxicology. Environmental Toxicology and Chemistry. 2013; 32: 1737–1739.

[22] Johnstone IM, Titterington DM. Statistical challenges of high-dimensional data. Philosophical Transactions of the Royal Society A. 2009; 367: 4237–4253.

[23] Kraemer HC, Stice E, Kazdin A, Offord D, Kupfer D. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. American Journal of Psychiatry. 2001; 158: 848–856.

[24] Zennaro F, Furlan E, Simeoni C, Torresan S, Aslan S, Critto A, *et al*. Exploring machine learning potential for climate change risk assessment. Earth-Science Reviews. 2021; 220: 103752.

[25] Byeon H. Determinants of blood pressure control in hypertensive individuals using histogram-based gradient boosting: findings from 1114 male workers in South Korea. Journal of Men's Health. 2024; 20: 47–55.

[26] Fang M, Shin M, Park K, Kim YS, Lee JW, Cho M. Analysis of urinary S-phenylmercapturic acid and trans, trans-muconic acid as exposure biomarkers of benzene in petrochemical and industrial areas of Korea. Scandinavian Journal of Work, Environment & Health. 2000; 26: 62–66.

[27] Capleton AC, Levy LS. An overview of occupational benzene exposures and occupational exposure limits in Europe and North America. Chemico-Biological Interactions. 2005; 153: 43–53.

[28] Gearhart-Serna LM, Tacam M III, Slotkin TA, Devi GR. Analysis of polycyclic aromatic hydrocarbon intake in the US adult population from NHANES 2005–2014 identifies vulnerable subpopulations, suggests interaction between tobacco smoke exposure and sociodemographic factors. Environmental Research. 2021; 201: 111614.

[29] Arnold SM, Angerer J, Boogaard PJ, Hughes MF, O'Lone RB, Robison SH, *et al*. The use of biomonitoring data in exposure and human health risk assessment: benzene case study. Critical Reviews in Toxicology. 2013; 43: 119–153.

[30] De Coster S, Van Leeuwen DM, Jennen DG, Koppen G, Den Hond E, Nelen V, *et al*. Gender-specific transcriptomic response to environmental exposure in Flemish adults. Environmental and Molecular Mutagenesis. 2013; 54: 574–588.

[31] Ephraim-Emmanuel BC, Ordinioha B. Exposure and public health effects of polycyclic aromatic hydrocarbon compounds in sub-Saharan Africa: a systematic review. International Journal of Toxicology. 2021; 40: 250–269.

[32] Kao TH, Chen S, Huang CW, Chen CJ, Chen BH. Occurrence and exposure to polycyclic aromatic hydrocarbons in kindling-free-charcoal grilled meat products in Taiwan. Food and Chemical Toxicology. 2014; 71: 149–158.

[33] Xu X, Liu X, Zhang J, Liang L, Wen C, Li Y, *et al*. Formation, migration, derivation, and generation mechanism of polycyclic aromatic hydrocarbons during frying. Food Chemistry. 2023; 425: 136485.

[34] Ledesma E, Rendueles M, Díaz MJ. Contamination of meat products during smoking by polycyclic aromatic hydrocarbons: processes and prevention. Food Control. 2016; 60: 64–87.

[35] Kumosani TA, Moselhy SS, Asseri AM, Asseri AH. Detection of polycyclic aromatic hydrocarbons in different types of processed foods. Toxicology and Industrial Health. 2013; 29: 300–304.

[36] Singh L, Agarwal T. Polycyclic aromatic hydrocarbons in diet: concern for public health. Trends in Food Science & Technology. 2018; 79: 160–170.

[37] Whaley CH, Galarneau E, Makar PA, Moran MD, Zhang J. How much does traffic contribute to benzene and polycyclic aromatic hydrocarbon air pollution? Results from a high-resolution North American air quality model centred on Toronto, Canada. Atmospheric Chemistry and Physics. 2020; 20: 2911–2925.

[38] Ali MU, Siyi L, Yousaf B, Abbas Q, Hameed R, Zheng C, *et al*. Emission sources and full spectrum of health impacts of black carbon associated polycyclic aromatic hydrocarbons (PAHs) in urban environment: a review. Critical Reviews in Environmental Science and Technology. 2021; 51: 857–896.

[39] Vardoulakis S, Giagloglou E, Steinle S, Davis A, Sleeuwenhoek A, Galea KS, *et al*. Indoor exposure to selected air pollutants in the home environment: a systematic review. International Journal of Environmental Research and Public Health. 2020; 17: 8972.

[40] Chen KC, Tsai SW, Shie RH, Zeng C, Yang HY. Indoor air pollution increases the risk of lung cancer. International Journal of Environmental Research and Public Health. 2022; 19: 1164.

[41] Yassin MF, Alhajeri NS, Kassem MA. Polycyclic aromatic hydrocarbons collected from indoor built environments on heating, ventilation and air conditioning dust filters. Indoor and Built Environment. 2016; 25: 137–150.

How to cite this article: Haewon Byeon. Identifying key risk factors of polycyclic aromatic hydrocarbons and benzene exposure in Korean adult males using machine learning approaches. Journal of Men's Health. 2025; 21(4): 26-33. doi: 10.22514/jomh.2025.049.