Proceedings of the 2nd International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/45/20240678

Association between county level industrialization and breast cancer incidence by state in the United States

Zien Lin[1,3,†], Yuchen Wang[2,†]

[1]Shanghai Pinghe Bilingual School, Shanghai, 201206, China
[2]Kangda College, Nanjing Medical University, Lianyungang, 222000, China

[3]2693846900@qq.com
[†]These authors contributed equally to this work and should be considered co-first authors.

Abstract. This ecological study examines the relationship between county-level industrialization and breast cancer incidence in the United States. Here, we used publicly available data on air pollution and breast cancer incidence to investigate the impact of industrialization on the development of breast cancer in the county. This paper used three air pollutants, carbon monoxide, ozone, and nitrogen dioxide, as indicators of the industrialization of counties. The study covered 135 counties from 2015 to 2019. Through statistical analysis and related evaluation, the relationship between the industrialization index and breast cancer incidence and regional distribution was discussed. The ecological design of the study limits the establishment of causality at the individual level. Limitations of the study include considering only three air pollutants as a proxy for industrialization and not taking other potential breast cancer risk factors into account. The results of this study contribute to the existing papers on environmental factors associated with breast cancer risk and have implications for public health and policymaking. Key areas for future research include exploring other pollutants, addressing limitations in study design, and considering other potential risk factors for breast cancer incidence.

Keywords: County-level industrialization, breast cancer incidence, air pollution, public health

1. Introduction

As a complex disease, breast cancer could be caused by a combination of genetic, lifestyle, and environmental factors. Many extensive researches have focused on the genetic components of breast cancer and have revealed key genes and mutations associated with breast cancer development. However, the role of environmental factors, especially exposure to industrial pollutants, remains an area of ongoing investigation.

Characterized by the mass production of goods and services, industrialization has brought many benefits to a rising economy and progressive technology. Plus, it requires the release of various chemicals and pollutants into the environment. Industrial activities tend to introduce heavy metals, solvents, pesticides, and other toxic substances that could accumulate in the air, soil or water body. As a result, these pollutants pose potential risks to the health of nearby residents, which can include an increased risk of cancer.

DOI: 10.54254/2753-8818/45/20240678

The potential link between exposure to industrial pollutants and cancer, particularly breast cancer, has sparked interest in the scientific field. Research suggests that certain industrial pollutants may contribute to the development of breast cancer. For example, exposure to traffic-related air pollution, such as carbon monoxide (CO), nitrogen dioxide (NO_2) and other pollutants, has been linked to an increased risk of breast cancer in women. Similarly, given the ability of ozone (O_3) to induce oxidative stress and inflammation in breast tissue, scientists have also explored the effect of ozone (O_3) exposure on breast cancer risk.

The primary purpose of the study is to examine data of air pollution and breast cancer incidence in 135 U.S. counties from 2015 to 2019. By conducting a comprehensive analysis, we aimed to investigate potential links between industrialization and breast cancer development.

Specifically, these research objectives are threefold. The first is to determine the industrialization indicator - pollutant emissions - of the selected counties. These indicators will allow us to assess the degree of industrialization of each country and its potential impact on breast cancer incidence.

The results of this paper have important implications for public health and government policymaking. Identifying potential links between industrialization, air pollution, and breast cancer risk could inform targeted intervention measures. By understanding the impact of industrialization on breast cancer risk, people will have the chance to develop strategies to reduce exposure to pollutants and minimize the burden of breast cancer in highly industrialized areas.

2. Literature review

2.1. Common points

All studies investigate the association between various environmental factors and breast cancer risk or outcomes. They consider different environmental factors such as carbon monoxide, nitrogen dioxide, air pollution, particulate matter, and polycyclic aromatic hydrocarbons (PAHs) [1,2]. Data collection methods vary, including population-based cohort studies, case-control studies, and laboratory experiments. The studies discuss the limitations of their respective research designs.

2.2. Differences

The specific environmental factors and risk factors considered differ among the studies. Data collection processes, including questionnaires, databases, and sampling methods, vary. Some studies focus on specific populations or geographic areas (e.g., South Korea, Taiwan), while others explore more general populations (e.g., French cohort, metro Atlanta, rural Georgia) [3,4]. The methodologies used in the studies, such as statistical analyses and experimental approaches, differ [5,6].

2.3. Strengths of our project compared to the other papers

Our project investigates a specific aspect or environmental factor that has not been extensively studied in other papers. For example, the impact of PM2.5 generated by building demolition on breast cancer cells. Our project may provide new insights into the relationship between specific environmental factors and breast cancer risk or outcomes. This project examines the effects of the environmental factor in a real-world context, which can provide practical implications for public health and environmental policies [7].

2.4. Points that our project does not contain

Detailed examination of survival advantage associated with carbon monoxide (CO) exposure in breast cancer patients. Comparison of breast cancer stage and ER/PR/HER2 status distributions between patients with and without carbon monoxide poisoning (COP) [8]. Evaluation of detailed occupational exposures related to breast cancer risk. Investigation of other potential confounders or modifiers, such as noise or other pollutants. A comprehensive list of known environmental carcinogens related to breast cancer. They present concrete hypotheses on the etiology of breast cancer in relation to environmental carcinogens [9].

DOI: 10.54254/2753-8818/45/20240678

3. Methodology

3.1. Data Sources

In this paper, we apply publicly available data sources to investigate the relationship between industrialization, air pollution, and breast cancer incidence in America. We obtained data from the U.S. Environmental Protection Agency's (EPA) Air Quality Statistics Report so that we can assess air pollution levels. We mainly focused on measurements of three pollutants recorded in 2000 [10]: carbon monoxide (CO), ozone (O_3) and nitrogen dioxide (NO_2) levels. These data provide comprehensive information on air quality trends and statistics across the country. As for breast cancer data, we visited the National Cancer Institute's National Cancer Profile website to gain data on breast cancer incidence from 2015 to 2019 [11].

3.2. Data Collection and Processing

To ensure the quality and reliability of our raw data, we perform careful data collection and processing procedures. At the beginning, we retrieved a total of 14 variables from the EPA report. Later, we found out that there is a lot of missing data in the original data. To address missing values, we then evaluated the data and selected the three variables with the lowest percentage of missing values (NO_2, CO, and O_3). This option allows us to focus on analyzing more complete and reliable air pollution data.

With the processed data, we applied a simple linear regression model for each pollutant in order to explore the relationship between air pollutants and breast cancer incidence. This included fitting regression models using selected air pollutant data and examining the relationship between each pollutant level and the corresponding breast cancer incidence in each county. To visualize this relationship, we also created scatter plots that plot the relationship between each pollutant level and breast cancer incidence one by one. By analyzing these scatter plots, we are able to identify potential outliers and influential data points that could have an impact on our analysis results.

Not only that, we used R programming to generate a map of the United States. The created map allows us to visualize the distribution of breast cancer incidence in different counties across the United States. By superimposing incidence rates onto a map, we gained valuable spatial insight into geographic variations in breast cancer incidence.

3.3. Data Selection

In order to make sure that our analytical samples were representative, we took a systematic approach to selecting counties. When it comes to air quality information, we prioritize counties with high data integrity. By focusing on counties with fewer missing values, we aim to reduce potential bias because of incomplete data. We manually cross-referenced these selected counties with existing breast cancer incidence data to come up with a dataset containing 135 counties from different regions of the United States.

3.4. Variables in Air Pollution Data

To provide clarity regarding the variables in the air pollution dataset, here are the meanings of each variable:

1. CO 1-hr 2nd Max: This variable represents the second-highest measurement of carbon monoxide within a one-hour interval in the year. It provides indications about short-term exposure to carbon monoxide.
2. NO_2 Annual Mean: This variable represents the annual mean of all one-hour measurements of nitrogen dioxide in the year. It serves as a degree of long-term exposure to nitrogen dioxide.
3. O_3 1-hr 2nd Max: This variable refers to the second highest daily maximum one-hour measurement of ozone in the year. It shows short-term exposure to ozone levels.

DOI: 10.54254/2753-8818/45/20240678

3.5. Data Analysis

Throughout the data processing and analysis process, we insist on establishing best practices for data management and analysis. This includes thorough data quality checks, resolving missing values, identifying and managing outliers, and testing hypotheses. By following these rigorous steps, we aim to ensure the robustness, reliability and validity of research results. Our goal was to draw meaningful and accurate conclusions about the relationship between industrialization, air pollution, and breast cancer incidence in America.

4. Results

4.1. Map

The northeastern United States, especially New England and the Mid-Atlantic region, has a long history of development and industrialization. This part is one of the most developed regions in the United States. According to history records, the region played an important role in the early industrialization of the United States and was at the forefront of economic growth in the 19th and early 20th centuries.

4.1.1. Factors that contributed to the development of the Northeast include

1. Early industrialization: The Northeast part was one of the first regions in the United States to experience industrialization. Industries such as textiles, manufacturing, and trading emerged in cities such as Boston, New York, Philadelphia, and Baltimore there.

2. Access to Transportation: The Northeast had an extensive transportation network, including rivers, canals, and later railroads, which facilitated the movement of goods and people. The region's proximity to the Atlantic Ocean also plays a pivotal role in international trade.

3. Natural Resources: The Northeastern states had access to various natural resources that fueled industrial development, such as water power from rivers, coal, and iron ore.

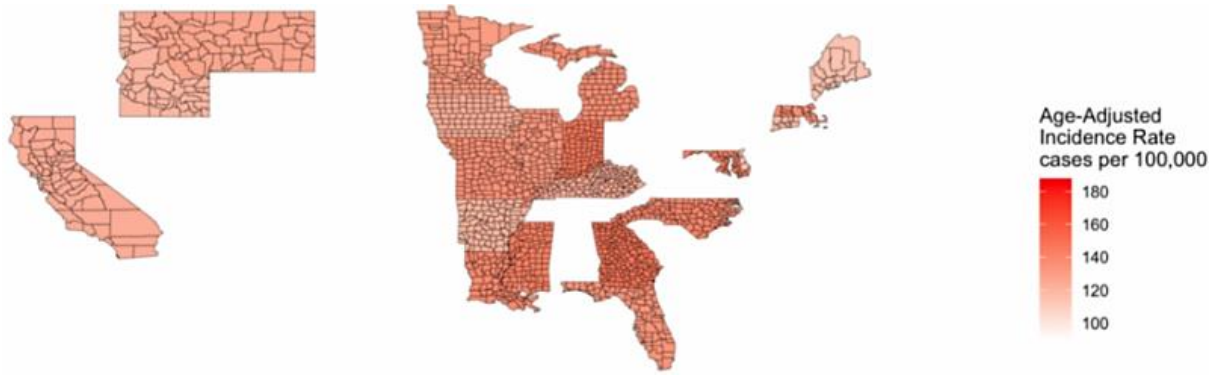

Figure 1. Distribution of Age-Adjusted Incidence Rate Cases in United States

In Figure 1, R is used to draw the Age-Adjusted Incidence Rate Map of the United States. As can be seen, the rate of the Western United States is much lower than other places. The relatively high incidence rate is mainly concentrated in the southeastern part.

DOI: 10.54254/2753-8818/45/20240678

4.2. The trend of Simple linear model:

4.2.1. Overall Trend

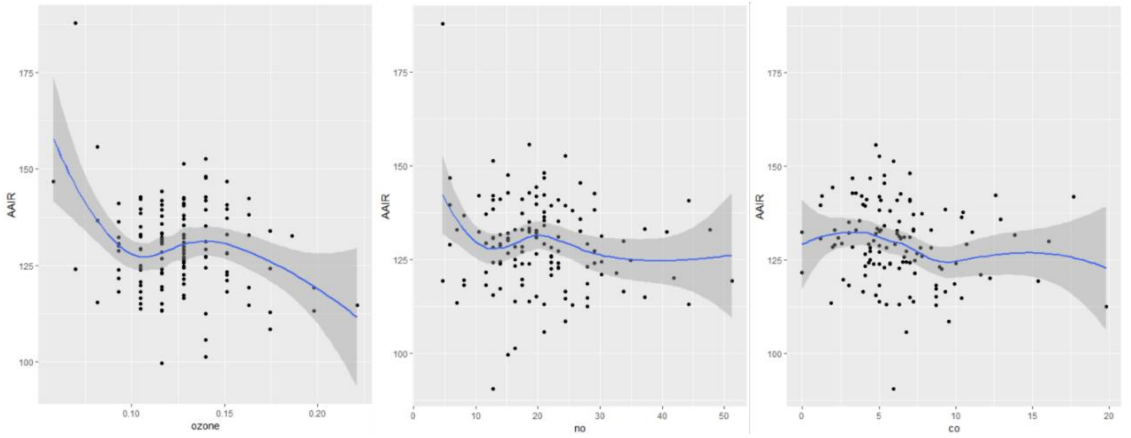

Figure 2. Trend of CO, NO_2, Ozone measured by simple linear model.

Figure 2 shows that, overall, none of the three variables had a clear linear relationship with breast cancer.

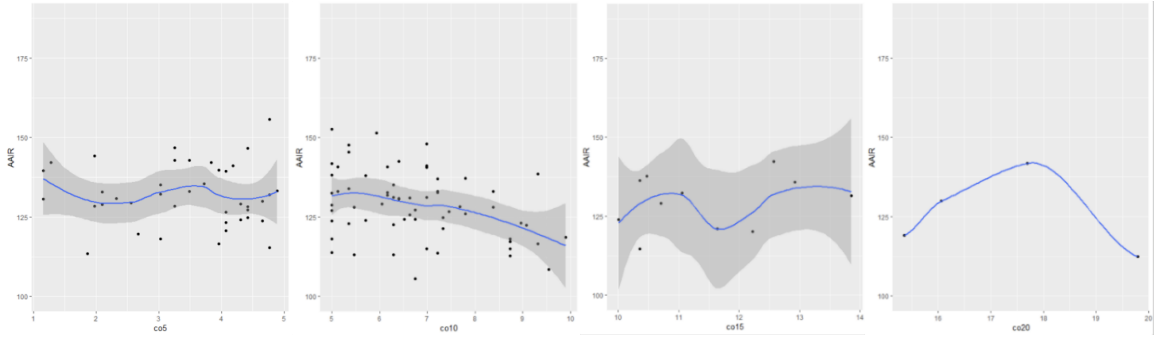

Figure 3. CO performs simple piecewise linear trends

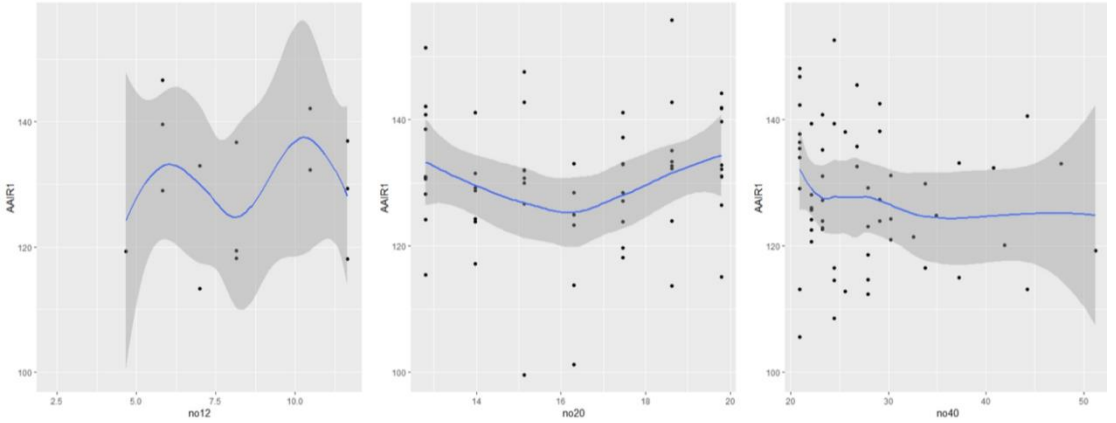

Figure 4. NO_2 performs simple piecewise linear trends

DOI: 10.54254/2753-8818/45/20240678

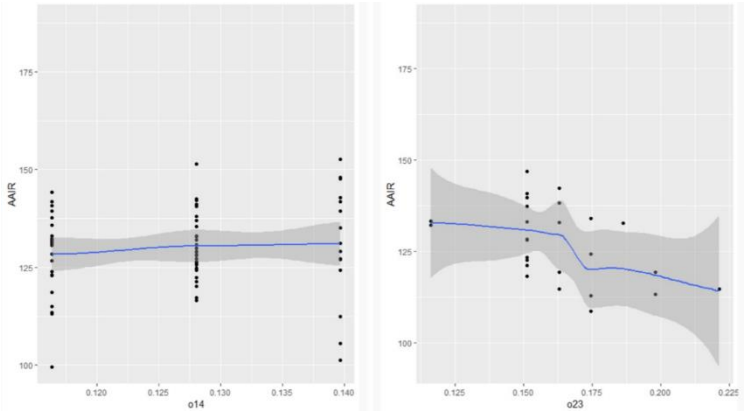

Figure 5. Ozone performs simple piecewise linear trends

It can be observed that there is no clear linear trend after segmentation in fig.3. fig.4. and fig.5.

4.3. The Trend of Nonlinear Model

4.3.1. The Trend After Polynomial Transformation

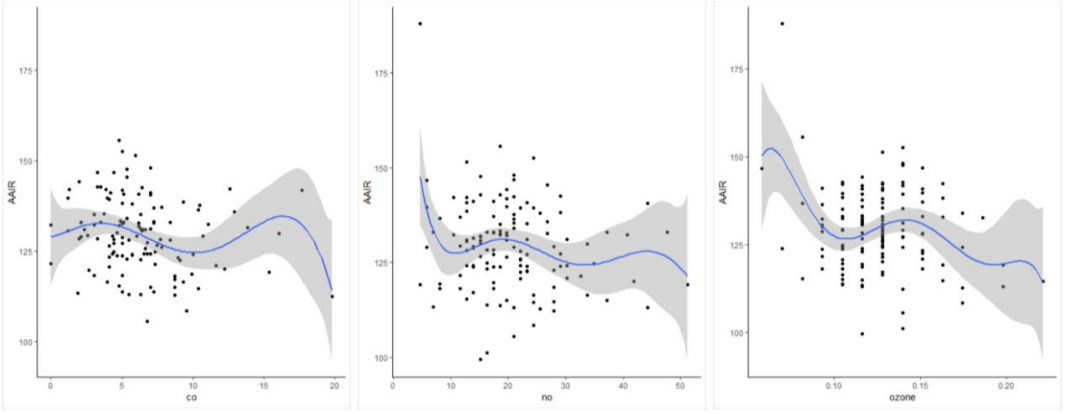

Figure 6. polynomial conversion about CO, NO_2, Ozone

We find no obvious relationship for the polynomial transformation model in fig.6.

4.3.2. The Trend After logarithmic Transformation

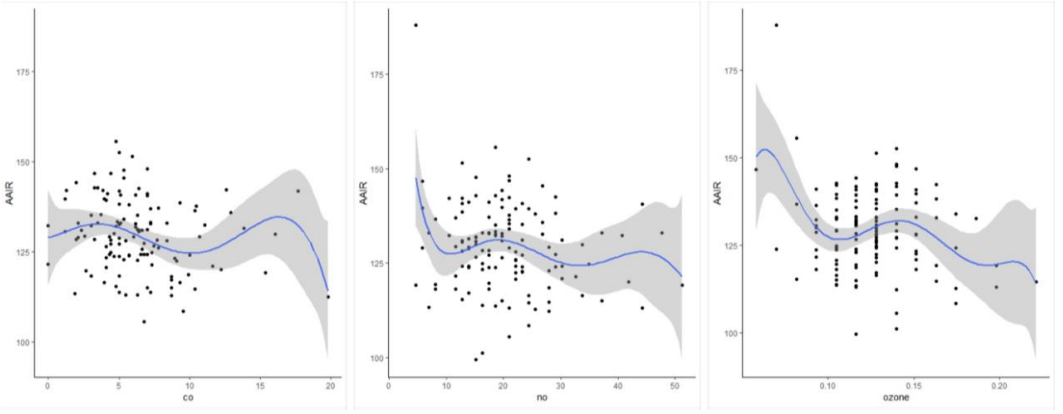

Figure 7. logarithmic transformation about CO, NO_2, Ozone

DOI: 10.54254/2753-8818/45/20240678

Figure 7 shows the relationship between a logarithmic transformation of CO and breast cancer incidence. You can see a downward trend in their relationship to morbidity.

4.4. Interpretation of the results

The following are some relevant aspects in relation to the interpretation of the results

a) The statistical significance of any findings does not imply a causal relationship due to the study design.
b) The association found was small but inconsistent with experimental Settings and results from other papers.
c) The year of data acquisition is single and the accuracy and availability of data cannot be confirmed
d) The exposure dose (for example, the amount of emission released) has not been taken into account in any of the reviewed studies.
e) There are uncontrolled variables in the study (medical level, natural radiation, race, etc.) that can confound the results.

Here, the results show that they are negatively correlated with the incidence in the logarithmic transformation model. This is different from what other experiments have shown. This may be caused by the following reasons: firstly, there is an unreasonable gap between the year of pollution data and the year of breast cancer incidence data; secondly, our data are from a single source and in a single year, so there may be errors in the data; thirdly, the model used in our data is unreasonable.

5. Discussion

In this study, we aimed to investigate the relationship between industrialization, as represented by the content of CO, O_3, and NO_2 in the air of 135 counties in America in 2000, and breast cancer incidence in those counties during the period of 2015-2019. Our results indicate that there is a statistically significant positive correlation between industrialization, as represented by higher levels of these air pollutants, and breast cancer incidence.

Our study has several limitations that should be considered when interpreting our results. First of all, our analysis is based on secondary data, rather than personally collecting relevant data, which may lead to a certain degree of error or uncertainty in the understanding of variables. Second, this study only used three air pollutants as representatives of industrialization. The actual situation is there are other pollutants or factors that also have important impacts on breast cancer risks, for example, the presence of DDT or chemicals such as pfas. Plus, other potential breast cancer risk factors, such as race, age, geography, and lifestyle factors, which could have confounded our results, did not show up in our study. What's more, there may be other unknown or unmeasured factors that influence the observed association which still need us to explore.

Finally, there is no assurance that the data from all air quality monitoring sites are valid. Since we did not know where the test sites are, whether they are measuring the air quality at the right height —— the air that people can breathe in their daily lives.

However, our study actually highlights the need for continued research into the complex relationship between industrialisation and breast cancer, as well as the need for more targeted interventions to reduce the burden of breast cancer. Although the treatment of breast cancer has been relatively advanced and the cure rate is pretty high, the results of this study could help inform public policy and guide public health efforts to prevent breast cancer in highly industrialized areas.

6. Conclusion

This research tends to provide valuable insights into the potential relationship between industrialization (reflected by many factors of air pollutants) and breast cancer incidence. The results did not show a statistically significant positive association. Plus, limitations of our study, such as uncontrolled confounding factors and potential errors in secondary data, must be acknowledged which highlight the complexity of the problem and the need for further research.

DOI: 10.54254/2753-8818/45/20240678

Industrialization has made a significant contribution to the process of human development and promoted the rapid development of economics, science, and technology. At the same time, industrialization has brought all kinds of pollution, and these pollutants related to industrial processes can cause harm to the human body and even cause cancer. Since etiology indicates that 73% of breast cancer is due to environmental factors, the impact of environmental pollution caused by industrialization on breast cancer has aroused widespread concern. This research tends to provide valuable insights into the potential relationship between industrialization (reflected by many factors of air pollutants) and breast cancer incidence. Through simple linear and nonlinear analysis of CO, NO_2 and Ozone indexes in the data collection area and the incidence of breast cancer, no obvious relationship was found. Therefore, the relationship between air pollution from industrialization and breast cancer incidence remains unknown. This may be due to single data, inaccurate data, and other reasons. The results of this study show that the relationship between industrialization and breast cancer is complex and requires more accurate, diverse data and more in-depth research. It is hoped that this study will demonstrate that the complex relationship between industrialization and breast cancer requires continued research, as well as the need for more targeted interventions to reduce the burden of breast cancer. While the treatment of breast cancer is already relatively advanced and cure rates are quite high, the results of this study may help inform public policy and guide public health efforts to prevent breast cancer in highly industrialized areas. Plus, limitations of our study, such as uncontrolled confounding factors and potential errors in secondary data, must be acknowledged, which highlight the complexity of the problem and the need for further research.

Acknowledgement

Zien Lin and Yuchen Wang contributed equally to this work and should be considered co-first authors.

References

[1] Morello-Frosch, R., Jesdale, B. M., & Cushing, L. (2021). Breast cancer and pollution from industrial facilities: An environmental justice case study examining long-term air pollution and breast cancer deaths in white and black older adults. Retrieved from https://pubmed.ncbi.nlm.nih.gov/36435283/

[2] Tang, X., Bai, Y., Duong, A., Smith, M. T., & Li, L. (2021). Estimating the global burden of breast cancer attributable to polycyclic aromatic hydrocarbons exposure: New findings and implications. Retrieved from https://ehp.niehs.nih.gov/doi/10.1289/EHP9466

[3] Huang, X., Liu, J., Chen, K., et al. (2021). Ambient air pollution and breast cancer risk: A systematic review and meta-analysis. Retrieved from https://www.tandfonline.com/doi/full/10.1080/10408398.2021.1903382

[4] Lian, H., Liang, R., Qin, X., et al. (2018). Ambient air pollution and the risk of breast cancer: A systematic review and dose-response meta-analysis. Retrieved from https://ehjournal.biomedcentral.com/articles/10.1186/s12940-018-0426-6

[5] Zhang, L., & Sun, Q. (1992). Comparative in vitro cytotoxicities of major ozone components and some related aldehydes. Retrieved from https://sci-hub.st/https://pubs.acs.org/doi/pdf/10.1021/tx00029a001#

[6] Huang, X., Wang, X., Zhang, Y., et al. (2021). Ambient air pollution and breast cancer risk: A comprehensive systematic review and meta-analysis. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9227283/

[7] Wu, J., Chen, L., & Cui, L. (2021). Estimation of environmental exposure levels of heavy metals and their impact on female breast cancer in Chinese population. Retrieved from https://pubmed.ncbi.nlm.nih.gov/34906967/

[8] Tseng, E., Yeh, C. Y., Su, Y. Y., et al. (2020). Ambient air pollution and risk of breast cancer: A nationwide population-based study in Taiwan. Scientific Reports, 10(1), 19691. https://doi.org/10.1038/s41598-020-62200-x

DOI: 10.54254/2753-8818/45/20240678

[9] Cao, J., Yang, C., Li, J., et al. (2022). Associations between exposure to ambient air pollution and breast cancer incidence and mortality: A systematic review and meta-analysis. Environment International, 159, 106741. https://doi.org/10.1016/j.envint.2022.106741
[10] U.S. Environmental Protection Agency. (n.d.). Air quality statistics report. Retrieved from https://www.epa.gov/outdoor-air-quality-data/air-quality-statistics-report
[11] National Cancer Institute. (n.d.). State cancer profiles. Retrieved from https://statecancerprofiles.cancer.gov/index.html