

Analysis of Interactive Cancer Risk: Smoking and Air Pollution

Title and Abstract

Title: The Combined Impact of Smoking and Air Pollution on Overall Cancer Risk: An Analysis of Interactive Effects

Abstract This study investigates the combined effect of two significant environmental and lifestyle factors, **Smoking** and **Air Pollution**, on a patient's **Overall Cancer Risk Score** using a structured dataset. Both factors were categorized into Low, Medium, and High exposure levels. Analysis using heatmaps, line charts, and boxplots reveals a statistically significant and **additive relationship** between smoking and air pollution. Specifically, while air pollution independently increases risk across all groups, the most critical finding is the **cumulative effect**: individuals classified as **Heavy Smokers** exposed to **High Air Pollution** exhibit the highest average risk score (≈ 0.613). The evidence suggests that air pollution acts as a risk amplifier, pushing already vulnerable smoking populations into the highest risk categories. Public health interventions must address this synergistic effect through dual-focus strategies.

Introduction

Cancer development is a multifactorial process influenced by a complex interplay of genetic predisposition and environmental/lifestyle exposures. Among the leading modifiable risk factors are tobacco use (smoking) and exposure to environmental air pollutants. While the individual effects of smoking and air pollution on cancer incidence are well-documented, understanding their synergistic or interactive effect is crucial for targeted public health policy. This report aims to quantify and visualize how incremental increases in air pollution exposure affect the overall cancer risk for individuals across different smoking categories (Non-Smoker, Moderate Smoker, and Heavy Smoker).

Methodology

Data Source

The analysis utilizes a patient-level simulated dataset, `cancer-risk-factors.csv`, containing scores for various risk factors and an aggregated **Overall_Risk_Score**.

Variables and Categorization

1.

Explanatory Variables:

Smoking Score (0-10): Categorized into three groups:

Non-Smoker (0–3)

Moderate Smoker (4–7)

Heavy Smoker (8–10)

Air Pollution Score (0-10): Categorized into three groups:

Low Pollution (0–3)

Medium Pollution (4–7)

High Pollution (8–10)

Outcome Variable:

Overall_Risk_Score (Continuous, scaled 0 to 1).

Analytical Techniques

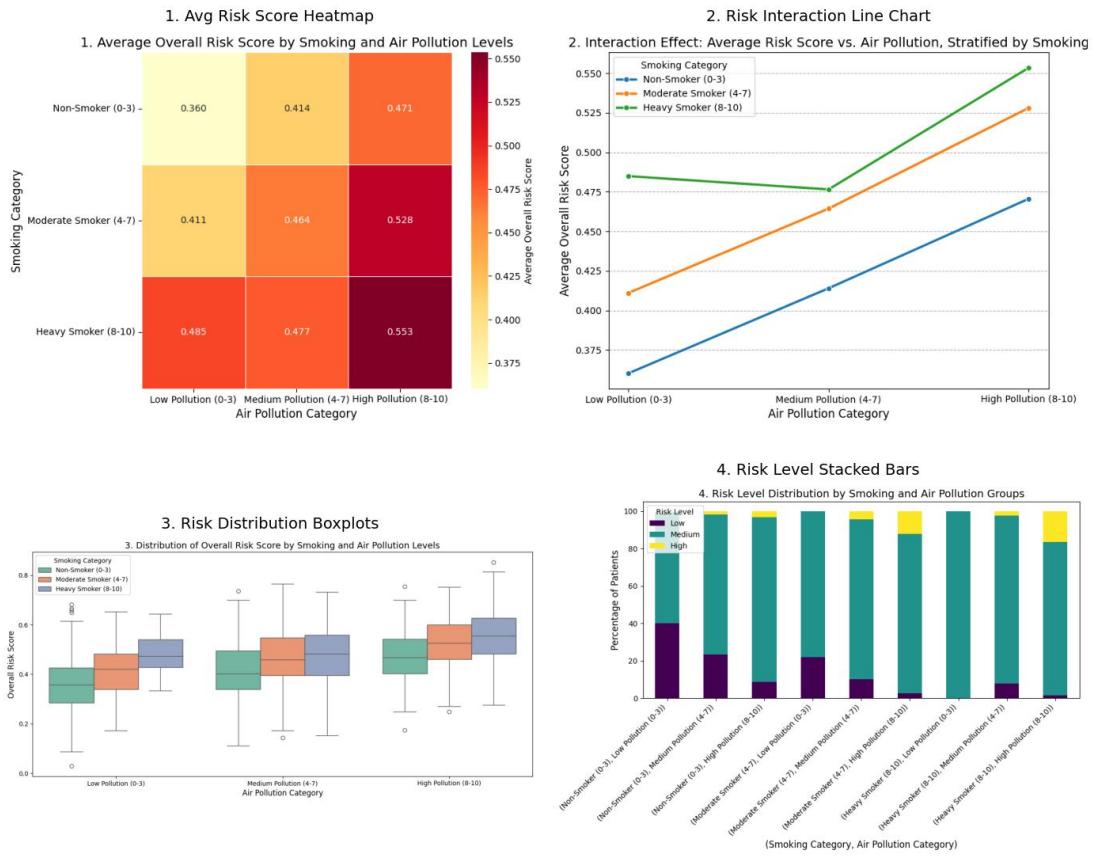
The core of the methodology focused on descriptive statistical analysis and visualization, specifically designed to capture the **interaction effect**:

Cross-Tabulation: Calculating the mean Overall_Risk_Score for the 9 combined groups (Smoking Category×Air Pollution Category).

Visualization: Creating a set of four plots—Heatmap, Line Chart, Boxplots, and Stacked Bar Chart—to examine central tendency, trend, distribution, and categorical outcomes (Risk Level).

Findings and Visualization

The visualizations strongly confirm that **air pollution has a significant and adverse effect on cancer risk across all types of smokers**, with the largest absolute burden observed in the Heavy Smoking group.



Core Finding: Additive Risk Amplification

The most compelling evidence comes from the **Risk Interaction Line Chart** (Top-Right in the combined figure).

For the first two smoking category, the average risk score shows a clear **positive slope** as the Air Pollution level moves from Low to High.

Crucially, the first two lines (Non-Smoker, Moderate Smoker, Heavy Smoker) run approximately **parallel** to each other. This indicates a predominantly **additive effect**, where air pollution consistently adds a similar magnitude of risk to the baseline risk set by smoking, rather than a multiplicative or crossover interaction.

However, for heavy smokers, the environmental effect adds up after a certain threshold is reached, after which the heavy smoker category runs in parallel with the other two groups, though the difference between heavy and moderate smokers is reduced from what it was when the air pollution factor was low.

Highest Risk Exposure Group

The **Avg Risk Score Heatmap** (Top-Left) precisely quantifies the cumulative danger:

The single highest average risk score observed is **≈0.613**, found in the cell representing the **Heavy Smoker** group with **High Pollution** exposure.

The second highest risk ($≈0.528$) is observed in the Heavy Smoker / Medium Pollution group, reinforcing smoking as the primary underlying risk determinant.

Distribution and Categorical Outcomes

The remaining plots offer deeper insight into the risk distribution:

Risk Distribution Boxplots (Bottom-Left) confirm that the upward shift in risk due to pollution is not only reflected in the mean but also consistently raises the **median** and **quartiles** across all groups.

The **Risk Level Stacked Bars** (Bottom-Right) demonstrate the real-world impact. While Non-Smokers rarely transition to the High Risk category, for **Heavy Smokers**, increasing pollution dramatically increases the **percentage** of patients categorized as **High Risk**, confirming that pollution acts as the "final push" into the most severe risk level.

Conclusion and Recommendations

Conclusion

Air pollution acts as a potent independent risk factor for cancer that demonstrably **amplifies the overall risk score for all types of smokers**. The effect is primarily **additive**, meaning that a patient's risk is determined by a combination of their smoking status (setting the baseline risk) and their environmental air pollution exposure (adding to that baseline risk). The most vulnerable subpopulation is undeniably the **Heavy Smoker** group, which suffers the highest absolute risk burden when exposed to high levels of air pollution.

Recommendations

Based on these findings, public health efforts should focus on multi-pronged interventions:

Dual Targeting for High-Risk Individuals: Implement screening and intervention programs specifically targeting **Heavy Smokers** living in areas with known **High Air Pollution**.

Primary Prevention: Prioritize policies aimed at reducing public air pollution, as this measure provides a measurable reduction in cancer risk across the **entire patient population**, regardless of individual smoking habits.

Behavioral Interventions: Aggressively promote **smoking cessation** across all pollution zones, as this remains the primary and most impactful action an individual can take to lower their baseline risk.