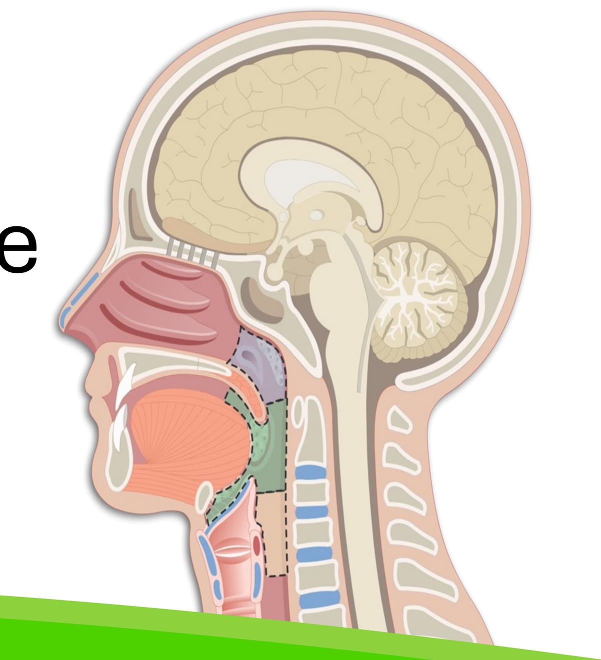
Air Pollution and Recurrence in HPV-Positive Oropharyngeal Squamous Cell Carcinoma: A Cohort Study on Long-term Exposure

Omnia Abdelrahman
Department of Epidemiology & Biostatistics
Boston University, Boston, MA
Email: Omnia@bu.edu

Robert Brody
Department of Otolaryngology: Head and Neck Surgery
Hospital of the University of Pennsylvania
Email: Robert.Brody@pennmedicine.upenn.edu



Abstract

Background: Air pollution is increasingly recognized as a factor in cancer outcomes. Patients with **HPV-positive oropharyngeal squamous cell carcinoma (OPSCC)** may be especially vulnerable to long-term exposure to pollutants such as **PM2.5**, **NO2**, and **VOCs** (volatile organic compounds, often found in emissions from vehicles, industrial processes, and indirect smoke from cigarettes, tobacco, marijuana, and vaping, such as benzene, formaldehyde, toluene, and acetone).

Objectives: This study aims to investigate the association between long-term air pollution exposure and recurrence in HPV-positive OPSCC patients.

Methods: A retrospective cohort of **791 HPV-positive OPSCC patients** was analyzed. Exposure to PM2.5, NO2, and VOCs was estimated using geocoded ZIP code data. Kaplan-Meier and Cox proportional hazards models were applied to assess the impact of pollutants on overall survival (OS) and recurrence-free survival (RFS), adjusting for confounders such as **smoking status**, **socioeconomic status (SES)**, and **comorbidities**.

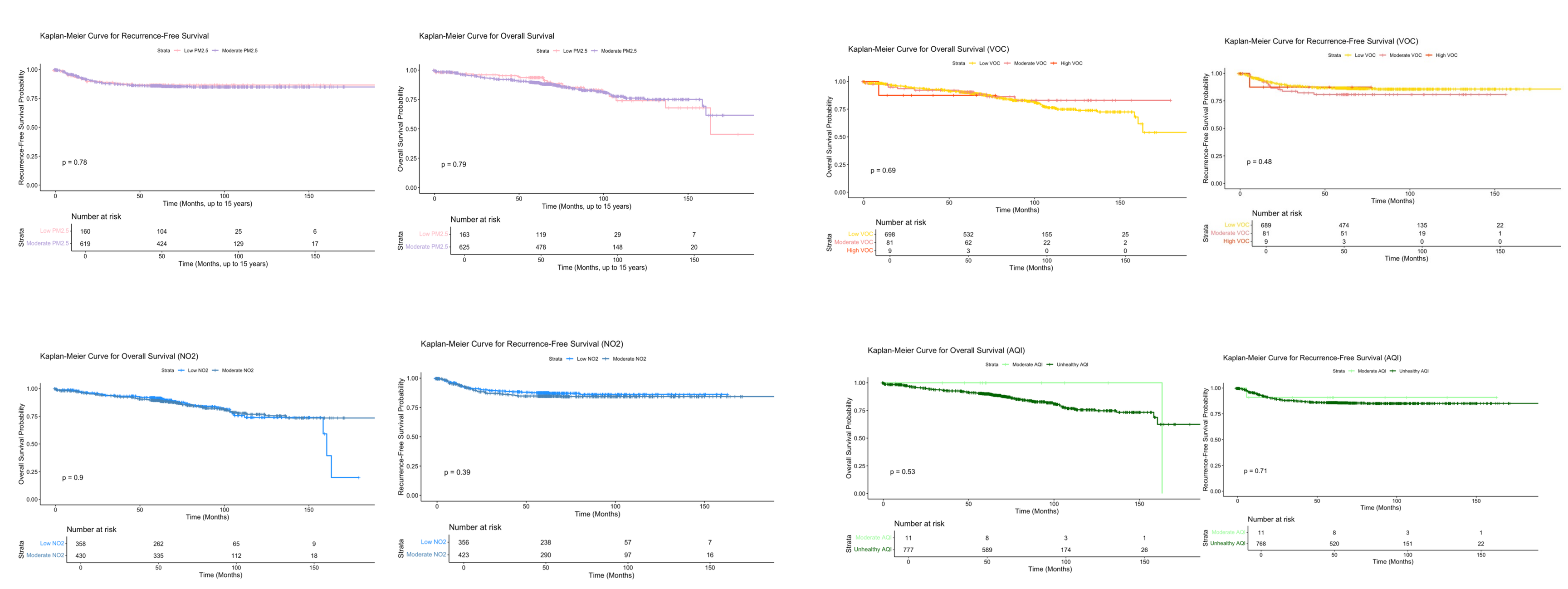
Results: Patients exposed to higher levels of **PM2.5** showed a **41.1% increased risk of recurrence** at 5 years (**HR: 1.411, 95% CI: 1.263 – 1.900, p < 0.0001**) and a **9.3% increased hazard of death (HR: 1.093, 95% CI: 1.001 – 1.192, p = 0.0466)**. Moderate **VOC exposure** was associated with a **26.3% increased recurrence risk** (**HR: 1.263, 95% CI: 1.001 – 1.592, p = 0.0466**), while NO2 exposure had no significant impact. **Socioeconomic status** further modified these risks, amplifying the effects of pollutant exposure.

Conclusion: Long-term exposure to **PM2.5** and **VOCs** significantly increases the risk of recurrence and decreases overall survival in **HPV-positive OPSCC patients**. These findings highlight the need for targeted **public health interventions** to mitigate environmental and socioeconomic disparities in cancer outcomes.

Introduction

Oropharyngeal squamous cell carcinoma (OPSCC) has been rising in incidence, largely due to the increasing number of cases driven by human papillomavirus (HPV) infection. Patients with HPV-positive OPSCC generally have a more favorable prognosis than their HPV-negative counterparts and often undergo transoral robotic surgery (TORS) as part of their treatment. Despite their overall better outcomes, these patients may still be susceptible to long-term environmental exposures, such as air pollution, which can influence cancer progression and recurrence in ways not yet fully understood.

Air pollution, particularly **fine particulate matter (PM2.5)**, **nitrogen dioxide (NO2)**, and **volatile organic compounds (VOCs)**, has been linked to several chronic health conditions. These pollutants tend to exert their effects through **long-term exposure**, making it challenging to detect their impact on cancer outcomes, such as recurrence or survival, in short-term studies.



Our study explores the relationship between chronic air pollution exposure and cancer outcomes in a specific population of **HPV-positive OPSCC patients treated with TORS**. While Kaplan-Meier survival curves did not demonstrate statistically significant relationships between pollution exposure and survival or recurrence, these findings may be influenced by factors such as the **complex nature of spatial variation** and the **uneven distribution of patients across exposure levels**. It is critical to adjust for confounders like **socioeconomic status**, **smoking history**, and **comorbidities** to accurately model the effect of environmental pollutants.

Understanding the impact of air pollution on this vulnerable population remains crucial, as even **subtle, long-term environmental factors** can contribute to cancer recurrence and affect quality of life. By addressing current limitations in biostatistical models and continuing research in this area, our study offers insights into the modifiable risk factors in cancer survivorship.

Methodology

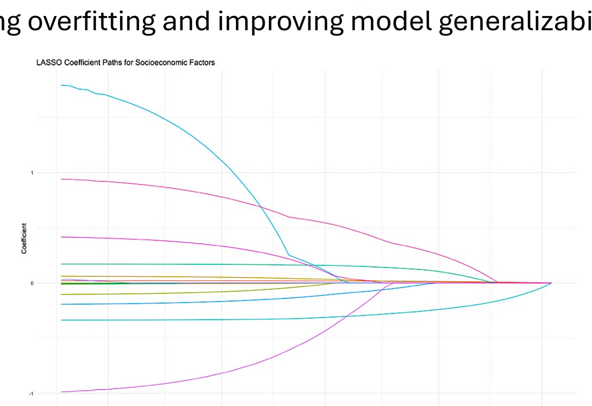
Our study's **biostatistical model** is distinct due to its focused approach, combining advanced statistics and spatial analysis to assess the impact of **chronic air pollution** exposure on cancer outcomes in **our cohort**. Below are the key methodological aspects:

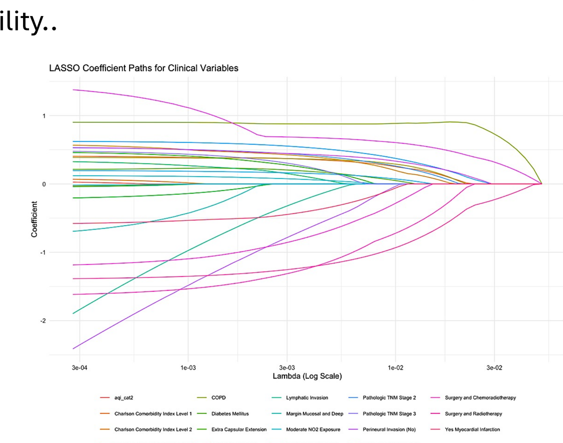
Data Collection & Exposure Assignment:

- Long-term exposure to **PM2.5**, **NO2**, and **VOCs** was assigned using **geocoded ZIP codes**, integrating spatial variation in exposure using CDC and EPA datasets (2007–2022).

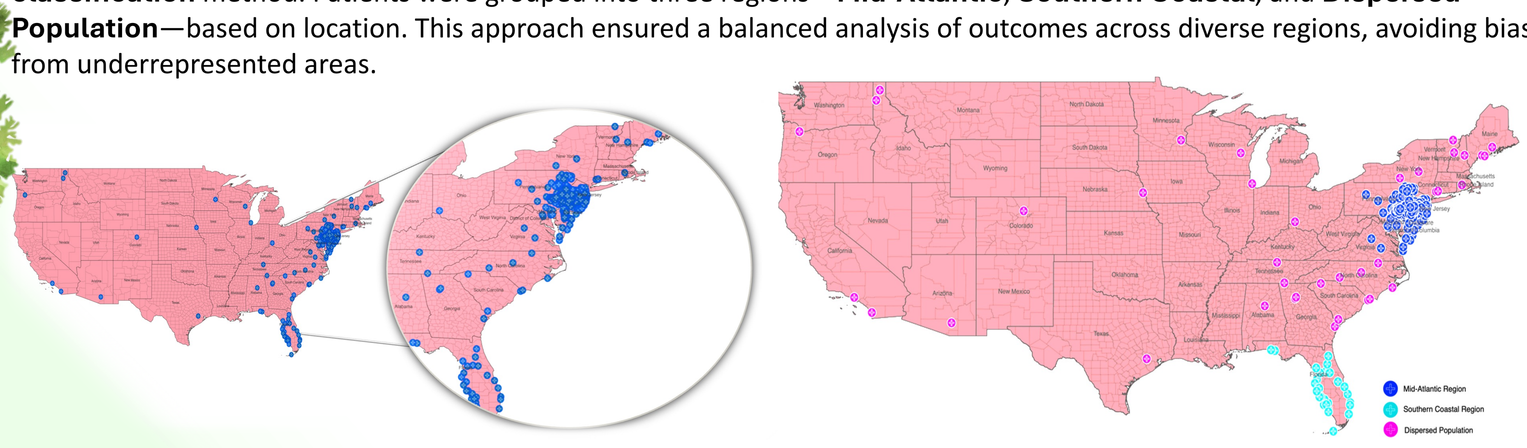
Environmental Metrics and Hierarchical Spatial Analysis:
A **hierarchical spatial approach** was used to address missing exposure data at the ZIP code level by assigning exposure based on **county then state-level** averages. This ensures that geographical variations in pollution are accurately reflected across the study population.

Statistical Model Design:
We used **LASSO regression** to select the most significant clinical and socioeconomic variables when adjust for all the 4 air pollution metrics used, reducing overfitting and improving model generalizability..

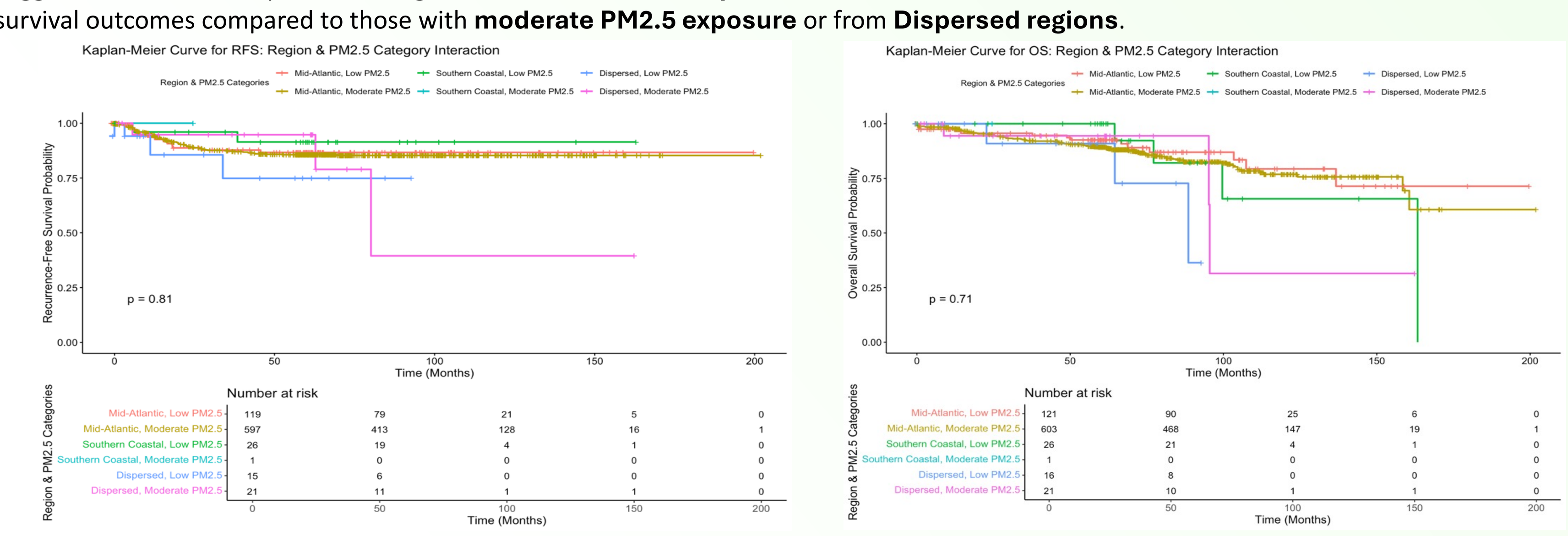




Geographic Classification of Patients Using Refined Hotspot Analysis to Avoid Dilution from UPenn Population
To prevent the dominance of UPenn's large patient population from skewing the results, we used a **refined geographic classification** method. Patients were grouped into three regions—**Mid-Atlantic**, **Southern Coastal**, and **Dispersed Population**—based on location. This approach ensured a balanced analysis of outcomes across diverse regions, avoiding bias from underrepresented areas.



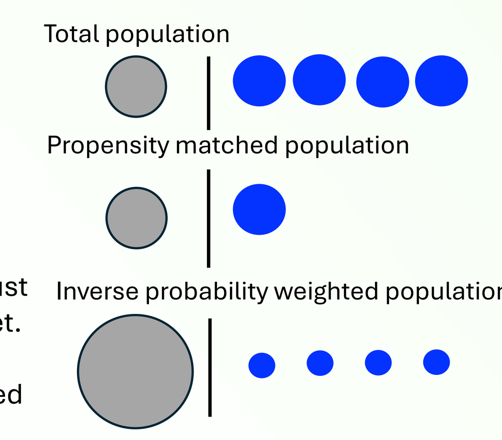
Kaplan-Meier Survival Analysis:
Kaplan-Meier curves were used to assess **Overall Survival (OS)** and **Recurrence-Free Survival (RFS)** across regions and **PM2.5 exposure levels** (low vs. moderate). Although not statistically significant (OS p = 0.71, RFS p = 0.81), the curves suggested trends where patients in regions with **lower PM2.5 exposure** and those from the **Mid-Atlantic** had better survival outcomes compared to those with **moderate PM2.5 exposure** or from **Dispersed regions**.



Estimating Propensity Scores with Ordinal Logistic Regression
We used **ordinal logistic regression** to estimate **propensity scores**, accounting for the key confounders that has been decided early on using LASSO model like **age**, **SES**, **smoking**, and **COPD** across regions. This step allowed us to balance the regions before conducting the weighted Cox proportional hazards model, ensuring a fair comparison between geographic areas.

Propensity Score Matching: Balancing Confounders
We applied **Propensity Score Matching (PSM)** to balance these confounders across regions, ensuring that differences in survival and recurrence outcomes are attributed to **PM2.5 exposure** and regional factors rather than underlying patient characteristics.

Weighted Cox Proportional Hazards Models
Following the propensity score matching, **Inverse Probability Weights** were applied to adjust for imbalances across the confounders (age, SES, smoking status, and COPD) in our dataset. The weighted Cox proportional hazards model was then used to estimate the association between **PM2.5 exposure** and survival outcomes. This approach minimized bias and allowed us to accurately assess the **hazard ratios** for mortality associated with PM2.5 exposure.

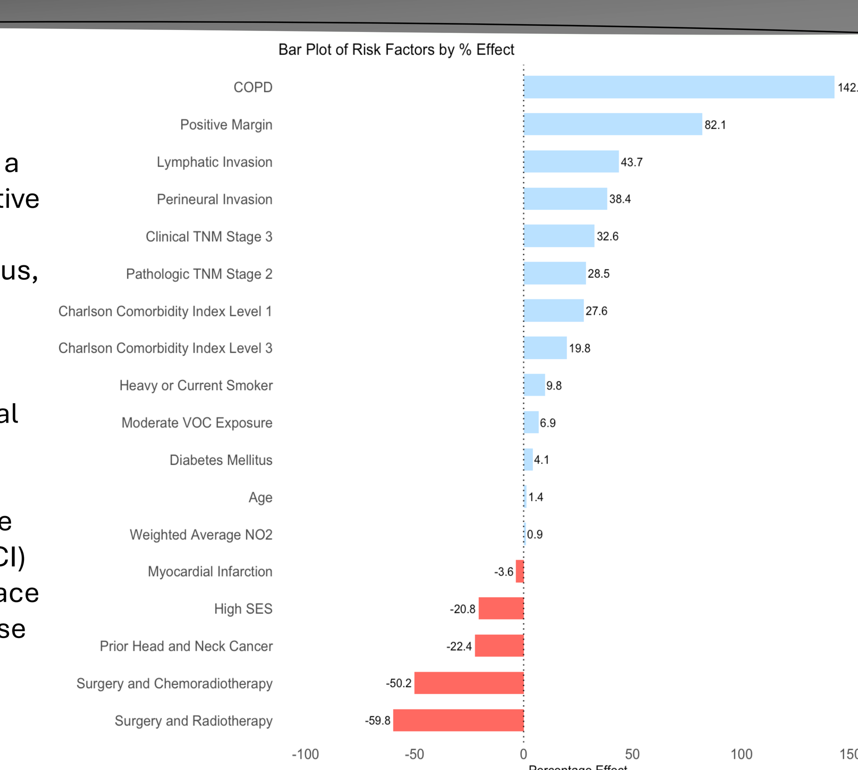


Results

LASSO Regression and Cox Proportional Hazards Models:

The LASSO regression results for **socioeconomic** and **environmental** variables show that patients with high socioeconomic status (SES) have a significantly lower risk of cancer recurrence (20.75% lower). This protective effect may be attributed to better access to healthcare and fewer environmental exposures. In contrast, factors such as age, smoking status, and exposure to air pollutants like nitrogen dioxide (NO2) and volatile organic compounds (VOC) increase the recurrence risk, with moderate VOC exposure raising the risk by 26.3%. These findings highlight the importance of both individual behavior (e.g., smoking) and environmental factors in cancer prognosis.

On the **clinical side**, key factors increasing the risk of recurrence include comorbid conditions like diabetes, high Charlson Comorbidity Index (CCI) scores, and advanced tumor stages. For example, patients with COPD face a 142.73% higher recurrence risk, while positive surgical margins increase recurrence risk by 82.1%. Conversely, undergoing surgery with radiotherapy or chemoradiotherapy substantially lowers the recurrence risk by 59.82% and 50.18%, respectively. These results underscore the significant role of clinical management and pre-existing conditions in predicting cancer recurrence.

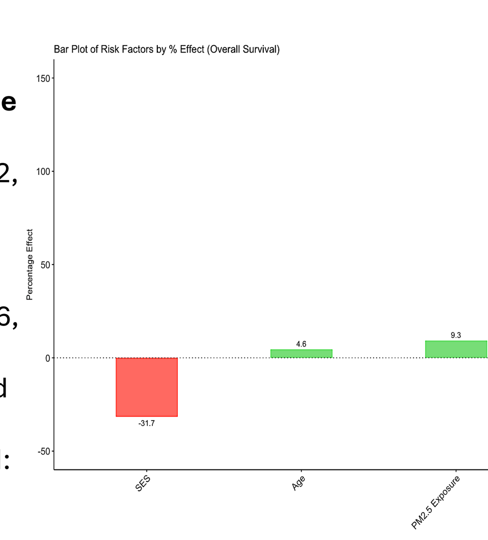


Propensity Score Matching (PSM) and Weighted Cox Models:

To address potential confounding, we applied **Propensity Score Matching (PSM)**, balancing key covariates such as **age**, **SES**, **smoking**, and **COPD** across regions. This allowed for a more accurate comparison of survival and recurrence outcomes by region and **PM2.5 exposure**. Inverse probability weights were generated from these scores and incorporated into the final weighted Cox models.

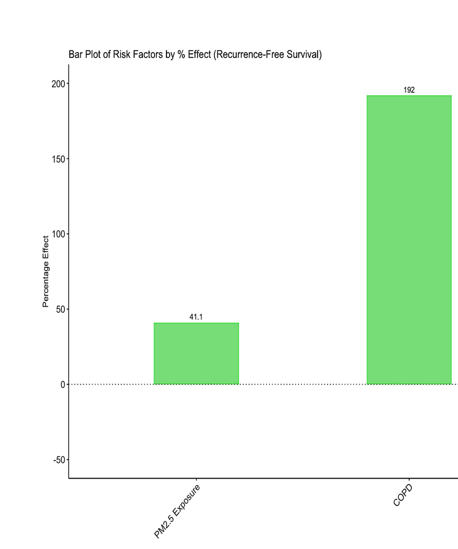
Overall Survival:

After adjusting for propensity scores, **PM2.5 exposure** was associated with a **9.3% increase in the hazard of death** (HR = 1.093, 95% CI: 1.001–1.192, p = 0.0466). Age remained significant, with a **4.6% increase in the hazard of death per year** (HR = 1.046, 95% CI: 1.043–1.049, p < 0.0001), while **SES** showed a **31.7% protective effect** (HR = 0.683, 95% CI: 0.638–0.732, p < 0.0001).



Recurrence-Free Survival:

PM2.5 exposure became significantly associated with recurrence at 5 years, with a **41.1% increase in the hazard of recurrence** (HR = 1.411, 95% CI: 1.263–1.573, p < 0.0001). **COPD** continued to be a dominant factor, contributing to a **192% increase in recurrence hazard** (HR = 2.922, 95% CI: 2.204–3.882, p < 0.0001).



Conclusion

Our study highlights a **complex interplay** between environmental exposures, socioeconomic status (SES), and clinical factors in determining cancer recurrence outcomes in HPV-positive OPSCC patients. Patients from lower SES backgrounds **not only** face higher cancer recurrence risks but also tend to reside in areas with poorer air quality, **compounding** their vulnerability. Contrarily, patients with higher SES enjoy both better access to healthcare and reside in cleaner environments, contributing to their lower recurrence risk. Furthermore, clinical factors such as smoking and comorbid conditions exacerbate these risks, particularly in low-SES populations. This highlights the need to address both **environmental and socioeconomic disparities** in **cancer care** to improve patient outcomes.

Public Health Message

HPV-positive OPSCC patients, while often having better prognoses than their HPV-negative counterparts, are **not immune** to worse outcomes—especially when exposed to long-term high levels of pollution. Our study highlights the urgent need to **follow up closely** with these patients, particularly those living in **lower socioeconomic areas with greater environmental risks**. Reducing their exposure to harmful pollutants like PM2.5 and VOCs must become a public health priority. By **advocating** for cleaner environments, **educating** patients about reducing exposure, and **ensuring** consistent long-term care, we can significantly improve their outcomes and protect them from avoidable cancer recurrences. These efforts are vital to ensuring that all patients, **regardless of their background**, have the opportunity for better health and survival.

Epidemiologists must account for the **clustering** of patients by socioeconomic and geographical factors to avoid **masking** key environmental and clinical risks. Patients living in poorer areas are disproportionately exposed to higher pollution levels and often face **barriers** to **accessing quality healthcare**, both of which elevate their risk for cancer recurrence. By integrating spatial and socioeconomic stratification into research and interventions, we can better understand and mitigate the compounded risks that these vulnerable populations face. Public health efforts should focus on reducing pollution exposure in these communities while improving access to care to ensure equitable health outcomes for all.