World Journal of Biomedical Sciences
ISSN: 2960-0189
DOI: https://doi.org/10.61784/wjbs3015

# GEO-SPATIAL ANALYSIS OF CANCER CLUSTER AND ENVIRONMENTAL RISK FACTOR IN THE USA

Md Mainul Islam*, Abdullah Al Mahmud Ashik, Safiul Islam, Shaid Hasan
*College of Graduate and Professional Studies, Trine University, USA.*
*Corresponding Author: Md Mainul Islam, Email: islammainul2019@gmail.com; mislam23@my.trine.edu*

**Abstract:** This study employs geo-spatial analysis to investigate cancer clusters and environmental risk factors across the United States (USA). Cancer is a significant public health concern, with millions diagnosed annually, and its etiology is multifactorial, involving genetic predispositions and environmental exposures. Environmental factors such as air pollution, proximity to hazardous sites, and lifestyle choices like smoking contribute to cancer incidence variations. Utilizing state and county-level cancer data from 2017 to 2021, this research employs GIS and statistical modeling to identify spatial patterns and clusters of cancer incidence. The findings highlight geographic disparities in cancer rates and underscore the complex interplay between environmental exposures and cancer risk, necessitating targeted public health interventions.
**Keywords:** GIS; USA; Cancer statistical map

## 1 INTRODUCTION

Cancer is one of the most prevalent diseases worldwide. Annually, 11 million individuals are diagnosed with different types of cancer. In the United States of America (USA), as of January 2022 over 18 million individuals had a history of invasive cancer. By 2024 it is projected that around 2 million new cases and 611,720 - 1600 cancer death per day- cancer deaths will occur in the country [1]. The burden of cancer is substantial, impacting individuals, families, and society as a whole. Numerous risk factors have been associated with various types of cancer, including age, race, ethnicity, socioeconomic status, behavioral factors and environmental exposures [2-4].

Environmental factors play a critical role in cancer development, encompassing a wide array of external influences as defined in epidemiology. These include lifestyle choices such as smoking, alcohol consumption, and dietary habits, all of which can significantly impact cancer risk. Radiation from both natural sources and environmental pollutants also contributes, highlighting risks associated with UV radiation and ionizing radiation, which is linked to secondary cancers post-treatment. Pathogens like HPV, hepatitis viruses, Epstein-Barr virus, and Helicobacter pylori are known contributors to various cancers, underscoring the role of infections in disease onset. Environmental carcinogenesis refers to the process where agents in our environment can cause cancer. The human body, while adapted to endure a certain level of exposure, faces significant cancer risk due to chronic exposure to known carcinogenic substances through air, water, food, and skin absorption. It's a highly complex, multi-stage process that unfolds over time and is not attributed to a single agent. This risk is particularly elevated in populations residing in industrialized regions or near factories. Carcinogens can alter DNA repair; induce epi-genetic alterations; cause oxidative stress; induce chronic inflammation; be immuno-suppressive; modulate receptor-mediated effects; cause immortalization; and alter cell proliferation, cell death, or nutrient supply. Genetic variations further complicate this picture, influencing individual susceptibility to these exposures. Conditions like Gilbert's syndrome signal an increased vulnerability to substances like acetaminophen. Importantly, the evidence underscores that cancer risk is not solely a matter of genes or environment, but rather a complex interplay between the two. The impact of common environmental exposures must be contextualized within the framework of underlying genetic susceptibility and windows of susceptibility, especially for early-onset cancers. The failure to consider these multifaceted factors has led to inconsistent and often contradictory findings in previous studies.

Despite the considerable toll of cancer on public health, there has also been some progress. The overall cancer deaths have dropped by 33% from their peak in 1991. This positive trend is largely attributable to advances in early detection and treatment, as well as lifestyle changes that have reduced exposure to certain risk factors [5].

However, the overall decline in cancer rates masks significant disparities across different population groups and geographic regions within the USA. Numerous studies have documented higher cancer incidence and mortality among certain racial/ethnic minorities, older adults, and socio-economically disadvantaged populations [6]. These disparities are thought to arise from differences in risk factor prevalence, access to screening and high-quality care, and tumor biology. In addition to demographic and socioeconomic disparities, significant geographic variation in cancer rates has also been observed within the US [7-8]. This geographic patterning likely reflects the uneven distribution of causal factors across different communities and regions, including environmental exposures, socioeconomic conditions, and health behaviors. Therefore, understanding these environmental and geographic disparities and their underlying determinants is crucial for developing targeted interventions and allocating resources effectively. However, evaluating environmental exposures and their relationship with cancer risk presents several intricate challenges. The transient nature of populations, exposure mixtures, dose-response relationships, and the timing of exposures (chronic versus acute) all contribute to the complexity of exposure assessment. For example, the long latency periods associated with many cancers, ranging from 8 to 40 years, make it challenging to establish causal links between environmental exposures and disease development.

Key considerations include statistical significance (whether the observed number of cases is higher than expected by chance), etiological relatedness of cases (whether the cases share a similar cause or etiology), well-defined and prolonged exposures, timing of exposures, and the ability to link the cluster to a specific environmental contaminant. Careful evaluation is necessary to distinguish true clusters from chance occurrences and to identify potential causal factors. To this end, geo spatial analysis (GIS) has emerged as a valuable tool in cancer epidemiology, allowing researchers to identify spatial patterns, clusters, and potential environmental risk factors for different cancer types [9-10]. Studies utilizing GIS have delineated the spatial distribution of various cancer types, including breast, colorectal, prostate, cervical, and childhood cancers [11]. These studies have explored associations between cancer incidence and mortality with environmental risk factors, such as air pollution, pesticides, and proximity to industrial facilities.

The primary aim of this study is to perform an extensive geo-spatial analysis of selected cancer types and their correlation with environmental risk factors in densely populated urban areas. Specifically, we aim to:

1. Investigate the relationships between cancer incidence rates and a wide range of environmental factors.
2. Identify areas exhibiting statistically significant clusters of higher or lower cancer incidence.
3. Determine whether the integration of spatial data and modeling techniques can improve the explanatory power of statistical models for cancer risk.
4. Link high-incidence outliers or hot-spots to potential contributing factors, such as proximity to environmental hazards or the presence of distinct population subgroups.

## 2 LITERATURE REVIEW

Geo-spatial analysis of cancer clusters and their environmental risk factors is an interdisciplinary area. Multiple sources of data and methods are combined to understand the spatio-temporal distribution of cancer incidence as well as mortality. Such an analysis is necessary for identifying cancer clusters and determining the environmental risk factors of these clusters, which can serve as the basis for the development of the public health measures and policies.

### 2.1 Cancer Clusters and Spatial Analysis

Greater than expected number of instances within an area during a given period identify cancer clusters. Understanding possible underlying causes—which can include environmental, occupational, or genetic elements—dependent on these clusters depends on their identification. Studies by Wheeler [12-13] have effectively utilized spatial cluster analysis techniques to pinpoint clusters of cancer incidence and mortality, with a particular focus on childhood leukemia and breast cancer. The importance of spatial analysis in identifying areas of high risk, which can help guide focused interventions and resource allocation for cancer prevention and control initiatives have been highlighted in these studies.

[14-15] found that while examining cancer clusters, it is critical to examine environmental risk variables. These findings emphasise the need to expand the study focus to incorporate lifestyle factors and environmental exposures beyond industrial pollutants to comprehensively analyze the causes of cancer. [16] examined the geo-graphical and temporal trends of bladder cancer mortality in the United States, identifying clusters of counties with a high prevalence of bladder cancer. This study highlights the significance of examining social, environmental, and occupational variables to gain a deeper comprehension of disparities in cancer outcomes based on gender, race, and geographical location.

### 2.2 Environmental Risk Factors

Environmental risk factors are crucial determinants in the etiology of cancer. Such factors include exposure to contaminants, emissions, agricultural chemicals, and lifestyle factors. Research has demonstrated a notable association between specific environmental exposures and an elevated likelihood of developing cancer. [17] investigated the link between cancer death rates and exposure to emissions from nuclear plants in the southeast of the United States. The study focused on the higher risks that come with these environmental factors. Studies by [18-19] also underline the need of qualitative spatial data analysis in determining the spatial distribution of malignancies and their frequency in different population. This research uses geographic information systems (GIS) and spatial statistical methods to clarify the causes of regional differences in cancer rates as well as the possible impact of outside variables on cancer risk. [20-21] study genetic and environmental variables in colorectal and prostate cancer. These studies show the complex interaction between inherited, family, and environmental factors in cancer etiology, with environmental influences affecting US cancer rates.

### 2.3 GIS and Cancer Research

GIS is needed to map cancer incidence and environmental concerns. Combining geographical data with cancer incidence data may help researchers find trends and links that epidemiological methods miss. [22] used GIS to study Long Island breast cancer cases. The researchers displayed cancer incidence and environmental exposure data, including industrial emissions and pesticide use, to identify risk areas. [23] research used genetic and geospatial analysis to pinpoint localized hotspots of diffuse gastric cancer in rural Central America. This work demonstrates the potential of geospatial analysis in identifying regional patterns of cancer risk factors by merging genetic, microbiological, and environmental data. These methods are essential for pin-pointing high-risk regions and developing focused actions to successfully reduce the cancer burden.

## 3 METHOD

Data included age-adjusted county-level cancer incidence for 3 cancers:

### 3.1 Study Region

The logistic regression modeling covered all states (n = 46) that had complete cancer population data. Among the original group of patients who had one of the four main types of cancer, multiple primary malignancies were identified in those who were diagnosed with at least one other type of cancer. County-level cancer incidence data were aggregated for the period spanning 2017 to 2021. The primary focus was on age and population-adjusted cancer incidence rates for all cancer types. To account for potential variations in spatial patterns across different geographic scales, the analysis was conducted at multiple spatial resolutions, including county, census tract, and zip code levels. This multi-scale approach allowed for a comprehensive understanding of the spatial distribution of cancer incidence and its potential associations with environmental factors.

### 3.2 Environmental Data

Various national and state-level data sources were utilized to obtain environmental data for correlation with cancer incidence rates. These sources included the United States census bureau, the centers for disease control and prevention (CDC)[1], the national center for health statistics (NCHS), and county health ranking. Additionally, databases maintained by the environmental protection agency (EPA)[2] and state level environmental monitoring programs were explored to incorporate comprehensive environmental exposure data, such as air pollution levels, proximity to hazardous waste sites, and potential exposure to environmental contaminants. The acquired environmental data were harmonized and geocoded to enable spatial analysis and integration with cancer incidence data. Environmental data were assessed using geographically weighted spatial regression analysis, adjusting for spatial autocorrelation.

### 3.3 Cancer Incidence Data

Cancer incidence data were obtained from two major nationwide cancer surveillance initiatives: the United States cancer statistics (USCS) program and the surveillance, epidemiology, and End results (SEER) program provided by the CDC[4], contains de-identified county-level[3]. The USCS, cancer incidence data reported to the CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's SEER registries. This integrated database encompasses data from all states in the U.S., the district of Columbia, and Puerto Rico, offering insights into over 33 million cancer cases. Permission was obtained from the SEER Program to access and utilize the data for this study. Stringent measures were taken to ensure the confidentiality and privacy of individuals, including avoiding any attempts to identify or contact patients, and not linking records to identifiable health information.

### 3.4 Geo-spatial Analysis

The Geo-spatial analysis was conducted in three main steps to discern the spatial distribution and patterns of age adjusted cancer incidence rates across U.S. counties within the study region.

#### *3.4.1 Exploratory spatial data analysis*

In the first step, the ASR were segmented into quintiles, and their spatial distribution was depicted using a choropleth map. This enabled the identification of unique spatial patterns and potential clusters of high or low cancer incidence rates.

#### *3.4.2 Global spatial auto-correlation analysis*

The second step involved spatial auto-correlation analysis, conducted using the Global Moran's I statistic, to examine the ASR across each county. The Global Moran's I statistic assesses both the locations and values of features to discern the existence of global spatial autocorrelation. In this analysis, the neighbor relationship was defined based on Queen's case adjacency, where all county boundaries sharing at least one corner were considered neighbors. To ensure the robustness of the findings, alternative spatial weight matrices, such as distance-based weights and higher-order adjacency, were explored. The Global Moran's I statistic for spatial autocorrelation is calculated as follows:

$$I = \frac{N}{s_0} \sum_i \sum_j w_{ij} \frac{(x_i - u)(x_j - u)}{\sum (x_i - u)^2} \tag{1}$$

- $N$ *is* the total number of counties in the study area.
- $s_0$ is a standardization constant based on the spatial weight matrix.
- $w_{ij}$ is the element in the spatial weight matrix corresponding to counties $i$ and $j$.
- $x_i$ and $x_j$ are the ASR for counties $i$ and $j$, respectively, with the mean $\mu$.
- $\mu$ is the mean ASR across all counties in the study area.
- The denominator $\sum(x_i - \mu)^2$ represents the varianceof the ASR across all counties.

***3.4.3 Local spatial autocorrelation analysis***

The Moran's I test detects overall spatial grouping, but it does not provide information about the exact locations of the clusters. The local indicators of spatial association (LISA) test, however, is employed to detect local clusters and calculate a statistic for each county. This computation utilises data from all counties to determine the statistical significance of the observed cancer rate in a particular county compared to neighbouring values. It is assumed that the rates conform to a Gaussian distribution. The LISA statistic is calculated by employing conditional permutation or bootstrapping techniques, which involve comparing the observed value with neighbouring values in a reference distribution. This aids in determining the magnitude of the value.

When assuming that spatial randomness is true, a result that is strongly connected with the neighboring values would be considered odd and would be found in the tail (rejection zone) of the test. The third step involved a local indicator of spatial association (LISA) analysis to examine the Anselin's local indicator of spatial auto correlation (ASR) for every county. LISA is a measure of local spatial association that determines whether the ASR of a particular county is more like its neighboring counties or to the overall average ASR across the entire study area. The significance of these associations was tested using a Monte Carlo permutation approach. The neighbor relationship was defined as Queen's case adjacency, where all county boundaries sharing at least one corner were considered neighbors. Clusters are considered statistically significant in the analysis when they have a P-value less than 0.05. Counties are categorised as either insignificant or belonging to one of four groups (high-high, low-low, high-low, and low-high) based on their cancer incidence rate compared to the average rate across all counties. The maps display the cluster's central point in colour, although the cluster's true size encompasses both the central point and its neighbouring points. The neighbours are initially determined via a matrix of weights assigned by a queen, and subsequently employed in the cluster analysis to establish the clusters. The cluster encompasses the neighbours, which are depicted as a grey buffer zone surrounding the centre. County boundaries have been omitted from all maps to comply with data restriction regulations. The spatial analyses were conducted using ArcGIS pro version 11.0.3.

***3.4.4 Regression modeling and proximity analysis***

Two approaches were used to determine how environmental factors affect cancer incidence. Gaussian geographically weighted regression (GWR) models were used to model ASR and environmental factors, taking geographical non-stationarity into consideration. An adaptive bi-square kernel function was utilised to compute spatial weights, with bandwidth optimised iteratively. Air pollution, proximity to hazardous waste sites, socioeconomic indicators, and land use/cover characteristics were considered environmental predictor variables.

The second investigation used GIS-based Proximity Analysis. We created Euclidean distance boundaries around hazardous waste sites and industrial facilities to start our analysis. These buffers measured counties' risk factor proximity. Additionally, network-based distance assessments measured the distance from county centroids to the nearest hazardous site along the transportation network. Last, kernel density estimation was used to detect environmental dangers and examine their impact on cancer rates. This comprehensive approach offered a solid framework for analysing the complex relationship between environmental factors and cancer rates.

***3.4.5 Sensitivity analysis***

Sensitivity analyses were conducted by varying parameters and assumptions of the spatial analyses, such as spatial scales, weight matrices, and model specifications. This approach helped assess the robustness of the findings and identify potential limitations or uncertainties in the analytical approach.

## 4 RESULTS

### 4.1 Sample statistics

The study included a large group of 6,533,532 persons who were diagnosed with primary cancer cases from 2016 to 2021. Out of the four main types of cancer examined, lung cancer had the largest share, representing 38% of cases. Skin cancer accounted for 36%, breast cancer for 24%, and brain cancer for 2%.

It is important to note that while environmental and carcinogenic factors contribute to the overall cancer burden, the study population comprised individuals from diverse demographic backgrounds, with 54.5% being female, 35.9% in the age group of 50-64 years, and 75.9% identifying as white.

***4.1.1 Lung cancer***

Lung cancer continues to be the leading cause of cancer related deaths among both men and women in US [24]. In 2024, it is estimated that there will be around 234,580 new instances of lung cancer and approximately 125,070 fatalities caused by lung cancer [25]. Epidemiological data indicate a strong correlation between lung cancer incidence and smoking rates. The highest incidence of new lung cancer cases are concentrated in the coal-mining and tobacco growing parts of the country. Kentucky having the highest incidence rate of 84.4 per 100,000 people, followed by West Virginia (75.7) and Arkansas (71.3). On the other hand, Utah (24.9) and Puerto Rico (15.1) had the lowest incidence rates. Incidence rates for lung cancer are generally falling across most states, with several states experiencing a significant declining trend, such as New Mexico (-5.0% annual percent change), Delaware (-5.8%), and Oregon (-3.6%). However, a few states like the District of Columbia and Minnesota showed a stable trend during the same period.

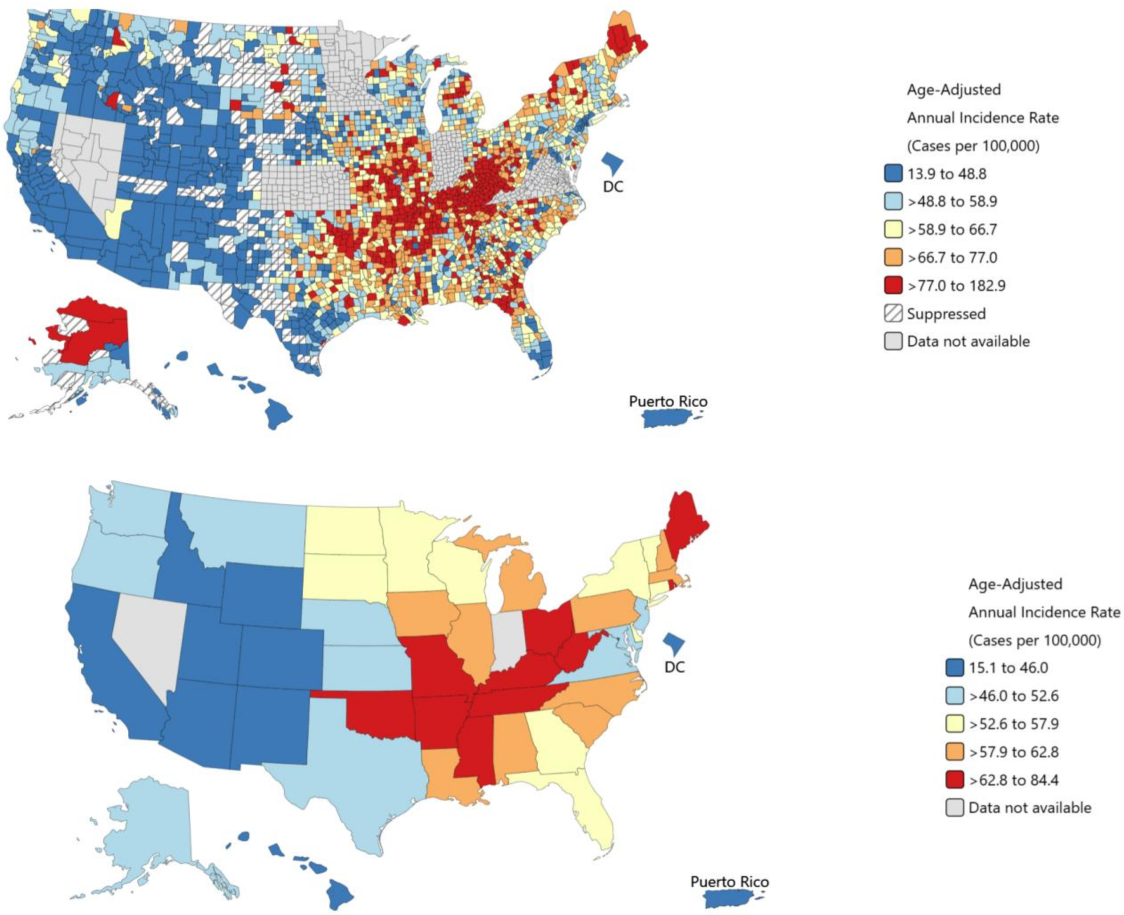


**Figure 1** Lung & Bronchus Cancer Incidence Rates United State by State, 2016-2021 (All Races, Both Sexes, All Ages)

The incidence rates and trends varied significantly across regions, with states in the Southeast, Midwest, and Appalachian regions generally showing higher rates compared to Western and Northeastern states. This pattern aligns with radon exposures, where Zone 1, covering parts of the Midwest and Appalachian region, exhibits the highest radon exposure levels. Zone 2, encompassing portions of the Great Plains and Midwest, follows with moderate exposure levels, while Zone 3, primarily comprising western and southeastern states, has the lowest potential for radon exposure.

Non-smokers, particularly women, exhibit higher rates of lung adenocarcinoma (LUAD), primarily attributable to air pollution [26]. Air pollutants, including particulate matter (PM10 and PM2.5), toxic metals, sulfur oxides, nitrogen oxides, and various microorganisms, contribute significantly to lung damage [27]. Despite a notable 24% reduction in fine particulate matter (PM2.5) from 2009 to 2016, levels have risen by 5.5% since, leading to approximately 10,000 premature deaths in 2018 [28]. This resurgence underscores the uneven distribution of air pollution across the U.S., with the West increasingly affected by wildfires, which exacerbate particulate matter spikes and degrade air quality in many cities. In 2022, an estimated 65 million individuals resided in counties with failing grades for short-term particle pollution. Conversely, while the East has historically contended with industrial pollution, it has experienced some improvements due to cleaner emissions from modernized industrial practices and vehicles. Nevertheless, disparities in air pollution exposure persist, disproportionately affecting people of color, who are 2.3 times more likely than white individualsto live in areas with poor air quality [26]. Figure 1 shows two maps of the United States depicting age-adjusted annual incidence rates of lung cancer at the state level (bottom map) and county level (top map).

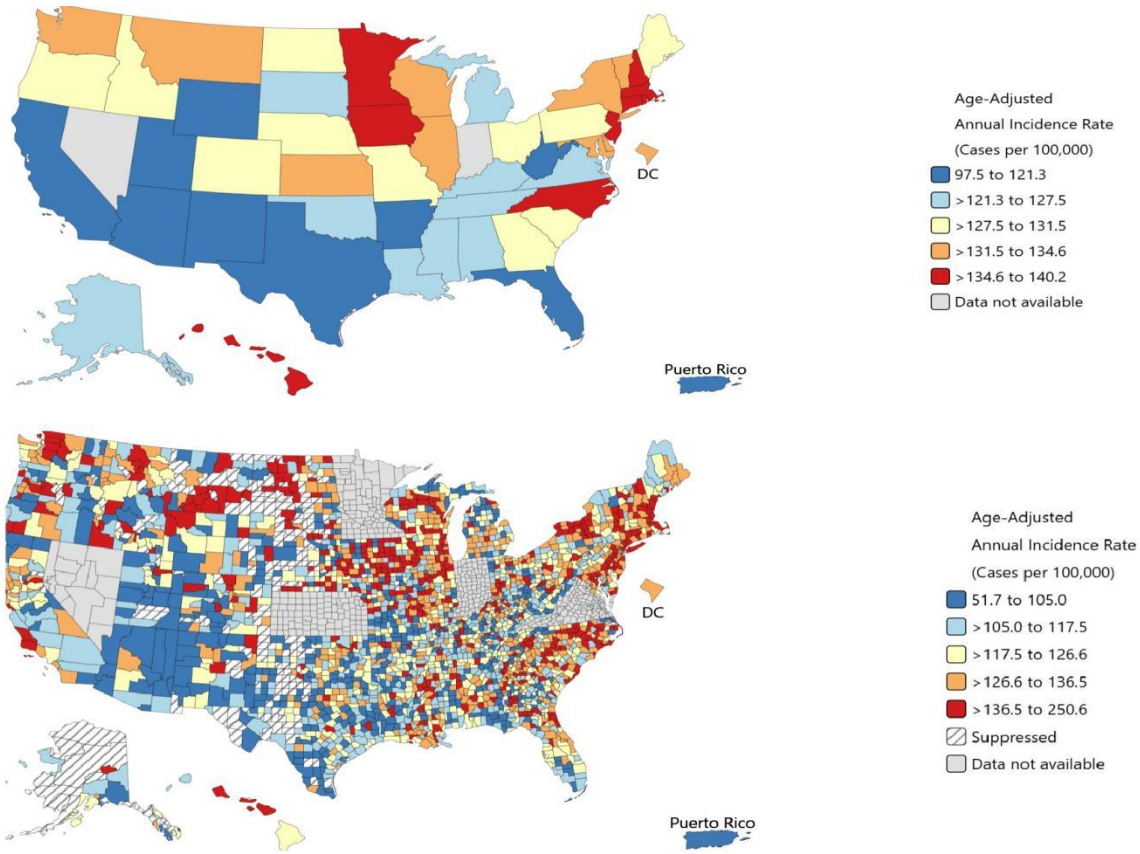


**Figure 2** Breast cancer Incidence Rates United States by State, 2016-2021 (All Races, Both Sexes, All Ages)

***4.1.2 Breast cancer***

Contrary to popular belief, only about 5-10% of breast cancers are attributable to high-risk inherited genetic mutations like BRCA1 and BRCA2. Studies of twins suggest that inherited genes account for approximately one third of breast cancer risk, leaving a substantial portion influenced by non-genetic factors, including environmental exposures [29]. Environmental chemicals can disrupt critical biological pathways implicated in breast carcinogenesis, including altering breast development, rendering the tissue more susceptible to Numerous common environmental chemicals have been linked to one or more of these biological pathways, despite not being adequately safety tested. These include endocrine disrupting compounds (EDCs) like bisphenol A (BPA), phthalates, parabens, and UV filters. Per- and polyfluoroalkyl substances (PFAS) used in stain-resistant products. Carcinogenic compounds like benzene, polycyclic aromatic hydrocarbons (PAHs), and pesticides.

The incidence rates for breast cancer display significant geographic variation across the US. Northeastern states like Massachusetts, Connecticut, New Hampshire, and Rhode Island have among the highest age-adjusted incidence rates, ranging from around 135 to 140 new cases per 100,000 women. In contrast, southwestern states like New Mexico, Arizona, and Utah tend to have the lowest incidence rates, mostly below 115 cases per 100,000. This geographic pattern aligns with known risk factors - the Northeast has higher rates of overweight/obesity and hormone replacement therapy use among postmenopausal women, while southwestern states have lower rates coupled with larger Hispanic populations who have relatively lower breast cancer risk. Looking at trends, most states are showing a slowly rising incidence, with lugares like Hawaii, Iowa, and Oregon exhibiting more rapid increases over the 2016-2021 period. However, several states including Maryland, Washington D.C., Vermont, and others displayed a stable trend with rates not significantly increasing or decreasing. Differences also emerged based on molecular sub types, with triple negative and basal-like breast cancers being more prevalent in African American women concentrated in the Southeast. This aligns with the higher overall incidence observed in this region compared to others. Figure 2 shows two maps of the United States depicting age-adjusted annual incidence rates of breast cancer at the state level (top map) and county level (bottom map).

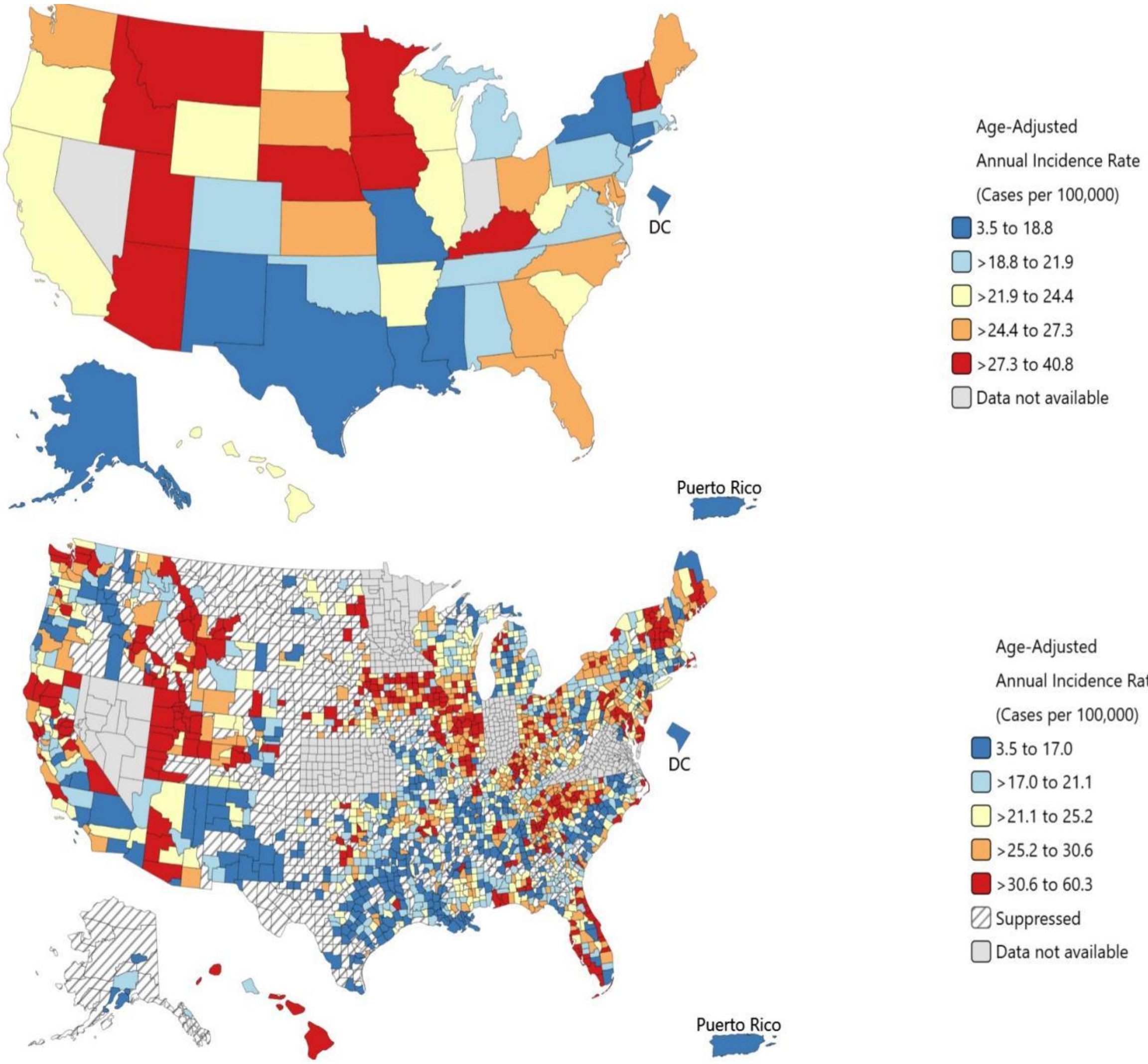


**Figure 3** Melanoma skin Incidence Rates United States by State, 2016-2021 (All Races, Both Sexes, All Ages)

### *4.1.3 Melanoma of skin*

The incidence rates of melanoma of the skin across the United States, measured between 2016 and 2021, reveal notable geographical variability and trends influenced by environmental factors such as UV exposure, altitude, and latitude. Utah reported the highest incidence rate of 40.8 per 100,000 people, likely driven by its high altitude and substantial UV radiation, while Puerto

Rico exhibited the lowest rate at 3.5, reflecting lower UV intensity and a potentially more protective demographic profile. States like Arizona (6.5), Illinois (4.8), and Texas (4.5), characterized by high UV exposure and outdoor lifestyles, showed rising melanoma trends, whereas others, such as South Carolina (-3.6) and Pennsylvania (-3.8), experienced declining trends possibly due to improved sun protection behaviors or effective public health campaigns. Meanwhile, Utah and New Hampshire maintained stable incidence rates despite their relatively high occurrence, indicating consistent environmental and behavioral risk factors over time. These rates were age-adjusted to the 2000 U.S. standard million population, with trends interpreted based on the Average Annual Percent Change (AAPC)and Annual Percent Change (APC). Stability in incidence was associated with a confidence interval of AAPC/APC that includes zero, while increasing or decreasing trends corresponded to intervals above or below zero, respectively. Larger confidence intervals, often due to lower case counts, indicate less stability in the data for certain states. Figure 3 shows two maps of the United States depicting age-adjusted annual incidence rates of melanoma skin at the state level (top map) and county level (bottom map).

## 5 CONCLUSION

In conclusion, this study underscores the critical role of environmental factors in shaping cancer incidence patterns across the USA. Through robust geo-spatial analysis, significant clusters of high cancer incidence were identified, often associated with environmental hazards such as air pollution and industrial sites. The research confirms previous findings of geographic variability in cancer rates and provides new insights into the spatial determinants of cancer risk. Moving forward, addressing these disparities requires integrated approaches that consider both environmental mitigation strategies and targeted healthcare interventions tailored to vulnerable populations. By leveraging advanced spatial analytics, this study contributes to the broader understanding of cancer epidemiology and informs strategies for reducing cancer burden through evidence-based policy and public health initiatives.

## COMPETINGINTERESTS

Theauthorshavenorelevantfinancialornon-financialintereststodisclose.

## REFERENCES

[1] Society AC. Cancer Facts & Figures 2024. American Cancer Society, Atlanta, 2024.
[2] Field RW, Withers BL. Occupational and environmental causes of lung cancer. Clinics in chest medicine, 2012, 33: 681-703.
[3] Filippini T, Heck JE, Malagoli C, et al. A review and meta-analysis of outdoor air pollution and risk of childhood leukemia. Journal of Environmental Science and Health, Part C, 2015, 33: 3666.
[4] Linder SH, Marko D, Sexton K. Cumulative cancer risk from air pollution in houston: disparities in risk burden and social disadvantage. 2008.
[5] Buck KD. Modelling of geographic cancer risk factor disparities in us counties. Applied geography, 2016, 75: 28-35.
[6] Ward E, Jemal A, Cokkinides V, et al. Cancer disparities by race/ethnicity and socioeconomic status. CA: a cancer journal for clinicians, 2004, 54: 78-93.
[7] Freeman VL, Ricardo AC, Campbell RT, et al. Association of census tract-level socioeconomic status with disparities in prostate cancer-specific survival. Cancer epidemiology, biomarkers & prevention, 2011, 20: 2150-2159.
[8] Schootman M, Jeffe DB, Lian M, et al. The role of poverty rate and racial distribution in the geographic clustering of breast cancer survival among older women: a geographic and multilevel analysis. American journal of epidemiology, 2009, 169: 554-561.
[9] Roquette R, Painho M, Nunes B. Spatial epidemiology of cancer: a review of data sources, methods and risk factors. Geospatial health, 2017, 12.
[10] Musa GJ, Chiang PH, Sylk T, et al. Use of gis mapping as a public health tool— from cholera to cancer. Health services insights, 2013, 6: HSI-S10471.
[11] Institute NC. Gis portal for cancer research. 2024. URL:https://gis.cancer.gov/.
[12] Wheeler DC. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in ohio, 19962003. International journal of health geographics, 2007, 6: 1-16.
[13] Amin RW, Fritsch BA, Retzloff JE. Spatial clusters of breast cancer mortality and incidence in the contiguous usa: 2000-2014. Journal of General Internal Medicine, 2019, 34: 412-419.
[14] Goodman M, Naiman JS, Goodman D, et al. Cancer clusters in the usa: what do the last twenty years of state and federal investigations tell us? Critical reviews in toxicology, 2012, 42: 474-490.
[15] Joseph N, Propper CR, Goebel M, et al. Investigation of relationships between the geospatial distribution of cancer incidence and estimated pesticide use in the us west. GeoHealth, 2022, 6: e2021GH000544.
[16] Smith ND, Prasad SM, Patel AR, et al. Bladder cancer mortality in the united states: a geographic and temporal analysis of socioeconomic and environmental factors. The Journal of urology, 2016, 195: 290-296.
[17] Wing S, Richardson D, Armstrong D, et al. A reevaluation of cancer incidence near the three mile island nuclear plant: the collision of evidence and assumptions. Environmental Health Perspectives, 1997, 105: 52-57.
[18] Babaee E, Roshandel G, Olfatifar M, et al. Geo-epidemiological reporting and spatial clustering of the 10 most prevalent cancers in iran. Geospatial health, 2021, 16.
[19] Ruktanonchai CW, Pindolia DK, Striley CW, et al. Utilizing spatial statistics to identify cancer hot spots: a surveillance strategy to inform community-engaged outreach efforts. International journal of health geographics, 2014, 13: 1-10.
[20] Ibrahem SQ, Al-Dalawi ZT, Bahaaldin AS. Sequence polymorphism in xenobiotic metabolising genes in iraqi colorectal cancer patients. Asian Pacific journal of cancer prevention: APJCP, 2021, 22: 1203.
[21] Fukagai T, Shimada M, Yoshida H, et al. Clinical-pathological comparison of clinical prostate cancer between japanese americans in hawaii and japanese living in japan. International journal of andrology, 2000, 23: 43-44.
[22] McMaster RB, Leitner H, Sheppard E. Gis-based environmental equity and risk assessment: methodological problems and prospects. Cartography and Geographic Information Systems, 1997, 24: 172-189.

[23] Dominguez RL, Cherry CB, Estevez-Ordonez D, et al. Geospatial analyses identify regional hot spots of diffuse gastric cancer in rural central america. BMC cancer, 2019, 19: 1-8.
[24] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA: a cancer journal for clinicians, 2022, 72.
[25] Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA: a cancer journal for clinicians, 2024, 74: 12-49.
[26] Liu X, Mubarik S, Wang F, et al. Lung cancer death attributable to long-term ambient particulate matter (pm2. 5) exposure in east asian countries during 1990-2019. Frontiers in medicine, 2021, 8: 742076.
[27] Manisalidis I, Stavropoulou E, Stavropoulos A, et al. Environmental and health impacts of air pollution: a review. Frontiers in public health, 2020, 8: 14.
[28] Thakrar SK, Balasubramanian S, Adams PJ, et al. Reducing mortality from air pollution in the united states by targeting specific emission sources. Environmental Science & Technology Letters, 2020, 7: 639-645.
[29] Cozen W, Lam E, Salehi M, et al. Contributions of twin studies to cancer epidemiology. In Twin Research for Everyone. Academic Press, 2022: 485-508.