Global Journal of Sciences
Volume 3, Issue 1, 2026, Pages 36-42
DOI: https://doi.org/10.48165/gjs.2026.3104; ISSN: 3049-0456
https://acspublisher.com/journals/index.php/gjs

Research Article

Spatial Analysis and Predictive Modeling of Cancer Incidence in Lake County-Illinois

Abdulrahman Abdullateef[1,*], Bello Sirajo[2], Mashood Olumuyiwa Lawal[3]
[1]*Department of Medical Biochemistry, Near East University, Nicosia, Turkey*
[2] *Department of Anatomy, College of Medicine and Health Sciences, Arabian Gulf University, Manama, Bahrain*
[3]*Department of Statistics, Faculty of Science, Air Force Institute of Technology, Kaduna, Nigeria.*

ARTICLE INFO

Article History
Received 15 Feb 2026
Revised 01 Apr 2026
Accepted 20 Apr 2026
Published 12 May 2026

Keywords
Bar plot;
Cancer;
Illinois;
lake county;
scatter plot;
spatial distribution

ABSTRACT

Certain regions may have higher cancer rates just by accident. Generally speaking, only when the pattern is statistically significant does it merit further investigation. This research was aimed at determining the relationships between cancer types, hotspots, and spatial patterns while considering environmental and demographic factors. An overview of the distribution of cancer and the relationships between different cancer types was provided by the data on cancer types gathered across ZIP codes from the Lake County Geographic Information System (GIS) in Illinois, as regards the cancer incidence updated (November 2024). Descriptive statistics such as minimum, maximum, and mean were used to summarize the data; skewness and outliers were visualized using a histogram. Scatterplots, bar plots, and spatial mapping were employed to describe the spatial variations and identified hotspots based on relative risks (RR). The relationships between cancer types were explored using correlation, and a linear regression model was employed to assess the factors that affect other types of cancer. Findings showed that cancers have significant maximum incidences such as breast, prostate, and lung cancers, while colorectal and urinary system cancers have lower minimum incidences. From the scatter and bar plot distribution of cancer incidence rates across the study, results revealed areas in Illinois with ZIP codes such as 60069, 60035, 60045 and 60044 are classified as hotspots, with higher chances of cancer incidence. Breast cancer and Urinary cancer System were significantly linked to the other types of cancers, through spatial mapping. These results highlight how critical it is to address demographic and environmental issues that might be causing these trends.

1. INTRODUCTION

A country's health is greatly impacted by cancer, lung cancer accounts for over one-fourth (23%) of cancer deaths, making it the second most common cause of death in the US. [1]. Meanwhile important advancements in the cancer care continuum have resulted in notable drops in cancer mortality rates over the past few decades, results also showed that younger persons are now more likely to be diagnosed with cancer. [2]. All cancer categories examined saw a decline in incidence rates among people 50 and older between 2011 and 2020, with the largest declines seen in male lung and bronchus cancer [3]. However, the incidence of cancer rose for female breast, female colorectal, and male colorectal cancer in people aged 20 to 49 [4]. According to estimates, the four most common malignancies diagnosed in the United States in 2022—lung, colon, breast, and prostate cancers—are also the major causes of cancer-related deaths [5].

*Corresponding author. Email: abdulrahmanabdullateef06@gmail.com

Prostate, female breast, and colorectal cancer deaths are also common. Compared to the US as a whole, Illinois citizens have greater mortality rates from all four of these cancers. [4].
These four major cancer types were predicted to cause 14,140 deaths in Illinois in 2022, with significant regional variation [6]. Compared to the US as a whole, Illinois citizens have greater rates of cancer death from all four cancer types. [7]. The Illinois Cancer Risk Index (ICRI), developed by earlier research, showed that population aging and increases in the size of the adult population were significant factors in the incremental number of cancer fatalities in the US [8].
The physical environment, clinical care, health behaviors, social and economic variables, and demographics are some of the elements that influence the occurrence of cancer [9]. For instance, Singh et al [10] at the individual and population levels, showed a correlation between socioeconomic status, race, and ethnicity with cancer outcomes in the United States. Cancer risk is influenced by cumulative environmental quality and health practices, as other researchers have documented [11]. In order to explain how social and biologic thinking, as well as dynamic and ecologic viewpoints, could influence distributions and lead to social disparities in cancer outcomes, an eco-social theory was reported and examined [2]. With regard to outcomes, the majority of cancer risk factors have been studied separately, however certain indexes have been developed and applied, but they only looked at one form of cancer [12]. However, our study covered a variety of cancer kinds, describing their hotspot and dispersion results that correspond to the four leading causes of cancer-related deaths in Illinois. [13].

2. PATIENTS AND METHODS

This is a population-based cross-sectional, observational, quantitative study, it included data from SEER registries licensed by the Lake County Geographic Information System (GIS), to examine the incidence of cancer in Lake County, Illinois. The northeastern part of the U.S. state of Illinois is home to Lake County, Illinois, a region with a variety of demographics and subdivisions at the ZIP code level. An overview of the distribution of cancer and the relationships between different cancer types was provided by the data on cancer types gathered across ZIP codes. Descriptive statistics such as minimum, maximum, and mean were used to summarize the data, skewness and outliers were visualized using histogram. Scatterplots, bar plots, and spatial mapping were employed to describe the spatial variations and identified hotspots based on relative risks (RR). The relationships between cancer types were explored using correlation, and a linear regression model was employed to assess the factors that affect other types of cancer.

3. RESULTS

The distribution of incidence rates for "Other Types of Cancers" throughout the study area is displayed by the histogram as shown in figure 1. Along the x-axis, incidence rates are grouped into bins, and on the y-axis, the frequency of ZIP codes within each bin is displayed. With incidence rates ranging from 800 to 1400 cases in the majority of ZIP codes, the data appears to be right-skewed. This skewness in the distribution indicates that cancer incidence is not evenly distributed, with outliers probably affecting regional averages. Only a few ZIP codes report occurrences that are abnormally high—nearly 2000 cases. Identifying ZIP codes with extreme values may help detect regional causes of high cancer rates. The level of relative risks (RR) is visually represented by a gradient color scale that runs from red (high danger) to blue (low risk). The RR and cancer hotspots across ZIP codes are displayed in the bar plot on the left panel of figures 3 and 4. With ZIP codes on the x-axis and RR values on the y-axis, the data is sorted by RR. Darker red tones indicate higher hazards, and the color gradient successfully illustrates the variation in cancer risk levels. The RR ranges from approximately 2400 to over 3300, suggesting significant variation

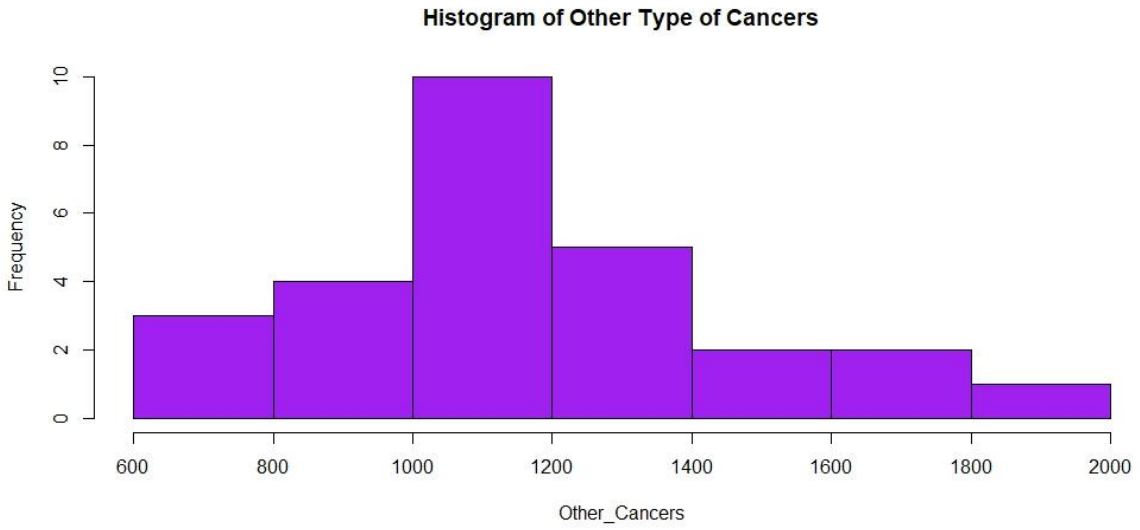


Fig. 1. Histogram of Other Types of Cancer

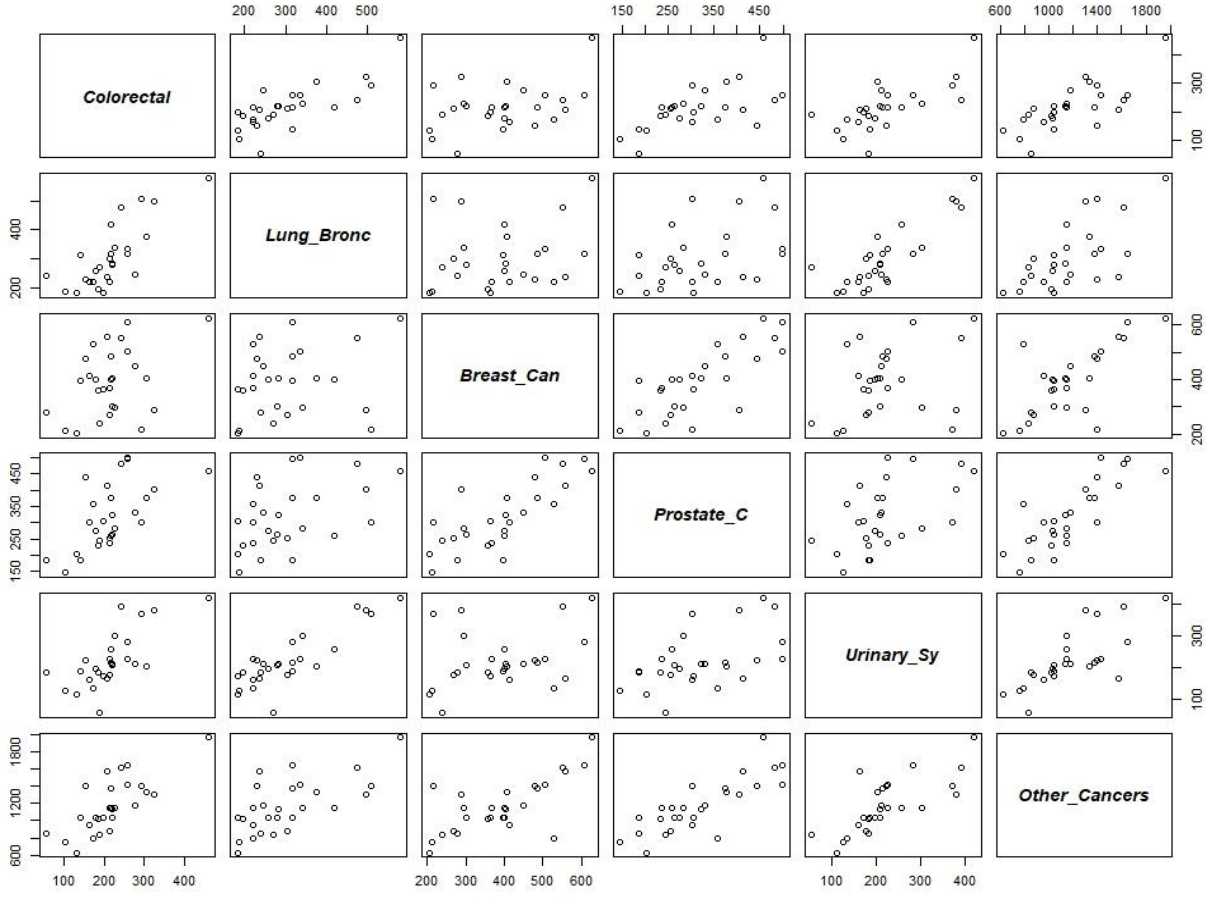


Fig. 2. Scatter Plot Between the Types of Cancer

in cancer risk across ZIP codes. A few ZIP codes exhibit particularly high RR (e.g., ZIP 60069, 60035, 60045), bringing attention to potential cancer hotspots. The right panel's bar plot groups ZIP codes into "Hotspot" and "Normal" groups according to whether or not their RR is greater than a predetermined threshold $(mean + 1 \times Std)$. The plot uses a binary color scale, with red denoting hotspots and teal denoting normal zones. The graphic shows a binary categorization rather than continuous data due to the constant bar height. The fact that most of them are "normal" ZIP codes suggests that high RRs are a regional occurrence. A few ZIP codes, including 60069, 60035, 60045, and 60044, are designated as hotspots, according to the left panel.

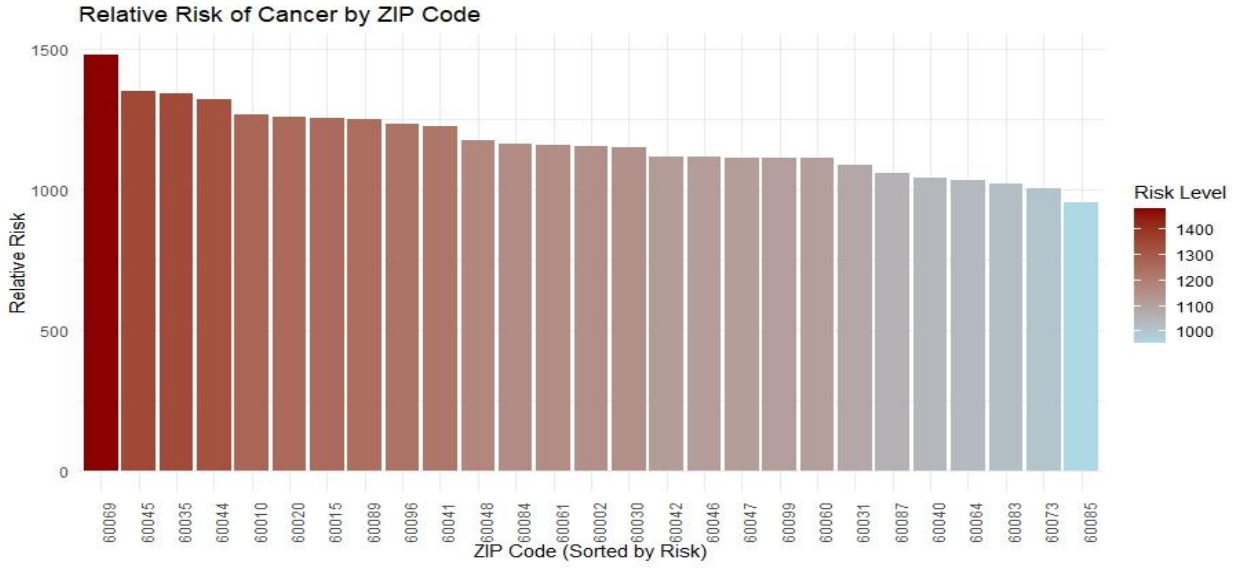


Fig. 3. Relative Risk of Cancer by ZIP Codes

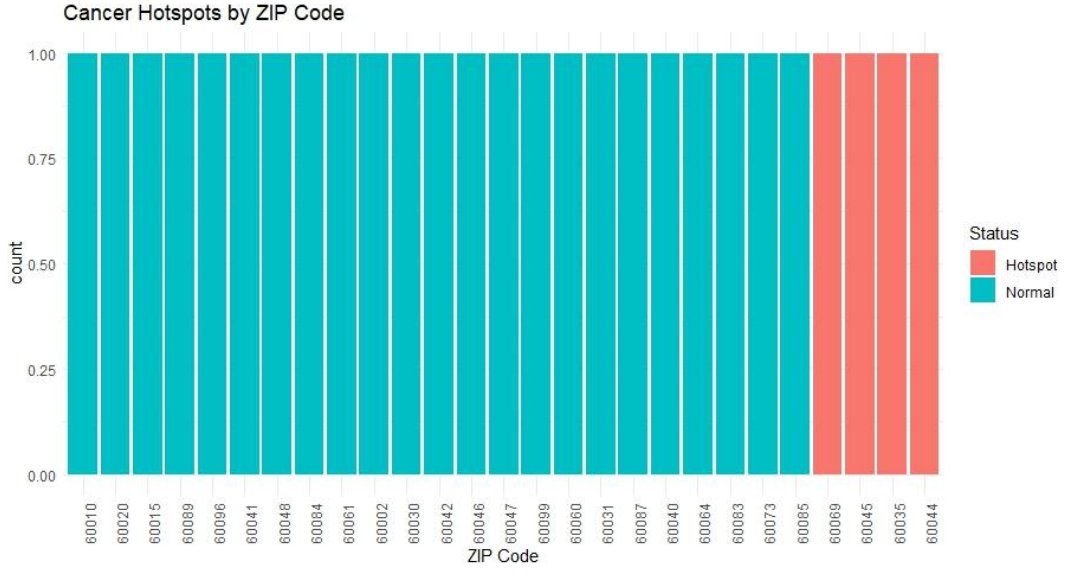


Fig.4. Hotspots of Cancer by ZIP Code

The spatial distribution of cancer incidence rates for "All Cancers" (left panel) and "Other Types of Cancers" (right panel) throughout the research region is depicted in the maps in Figure 5. Regional variations are emphasized with a color gradient that ranges from pale green (low incidence) to dark red (high incidence). In the case of "All Cancers," certain regions, like the southeast, have the highest incidence rates. The northern regions, however, report comparatively low incidences. Although the range of incidence values is smaller and the northern regions seem more uniform, the "Other Types of Cancers" map displays a similar spatial pattern. Patterns for "Other Types of Cancers" generally match those for "All Cancers," according to a comparison of the two maps, suggesting that these types play a major role in the overall incidence of cancer.

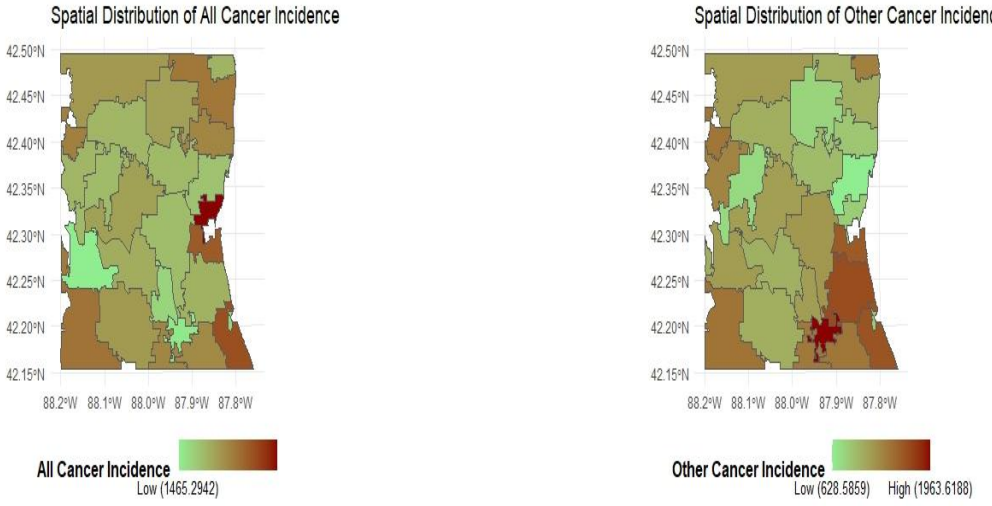


Fig. 5. Spatial Distribution of Cancer Incidence

To evaluate the factors influencing the response variable (other types of cancers), the linear regression model incorporates predictors such as colorectal, lung, breast, prostate, and urinary system cancers. With an adjusted R-squared value of 0.8316, the model's overall fit is strong, indicating that the included predictors can account for about 83% of the variability in other cancer types. Unaccounted cancers, or other types of cancers, were found to be significantly predicted by cancers of the breast and urinary system. Higher incidences of unclassified or unaccounted-for cancers are linked to higher rates of breast cancer, according to a statistically significant positive correlation between breast cancer and other cancer types (estimate = 0.8545; p-value = 0.0202). At the 5% significance level, urinary system cancer was also significant (estimate = 1.2533; p-value = 0.0458), indicating that areas with higher rates of urinary system cancer also have higher rates of other cancers. Despite their estimates suggesting possible associations, other predictors, including colorectal, lung and bronchus, and prostate cancers, did not achieve statistical significance ($p > 0.05$).

4. DISCUSSIONS

We constructed a novel relative risk and hotspot incidence of cancer by ZIP code by using population-based, Lake County data from the state of Illinois. The ZIP code cancer incidence represented a broad range of level in cancer diseases and their types. Our study differs from other reports that also describe cancer risk indexes. Although Wang et al [14] used data from 1998–2000 to focus exclusively on late-stage cancers and did not include values for different counties. Meanwhile our study involves data from

2022-2024. Gou et al [9] had more counties in Illinois classified as high risk, while our study observed a lesser high-risk classification in the counties at Illinois[9].

According to the descriptive statistics reported, while some types of cancers have significant maximum incidences such as breast, prostate, and lung cancers; others have lower minimum incidences such as colorectal and urinary system cancers. This suggests that there may be hotspots or areas that require focus as in the case of Qiu et al [13]. The high cancer cases, as indicated by the average incidence of 320.1, is in conformation with national trends, which depict that prostate cancer rates are higher in some certain demographic groups. The average incidence of colorectal cancer is 215.66, ranging from 55.56 to 457.78. This variance suggests that the incidence of colorectal cancer varies significantly among ZIP codes. The breast cancer cases vary significantly by region, with a mean of 392.8 and a maximum of 626.4, these differences may be due to easy-access to screening or demographic factors. Lung and bronchus cancer is still a major health concern in the area, as indicated by the mean incidence of 305.6, which ranges from 182.8 to 578.2.

Our results imply that regions with high rates of one among the common types of cancer, may also have high rates of another type amongst the common types, possibly as a result of common risk factors like environmental exposures or lifestyle. It is worth noting that each cancer type has a strong correlation with all cancers in comparison with the other cancers, which avoids multicollinearity concerns in the subsequent regression analysis. The framework in our study could potentially be used to create a similar scoring system for public health professionals.

5. CONCLUSION

This article introduces the Illinois the ZIP code cancer incidence, which incorporates the occurrence of the most common cancer incidence within the lake County. The index can be used to identify counties in Illinois where public health intervention is needed most. Particular ZIP codes are crucial hotspots for monitoring cancer incidence, highlighting the necessity for precise public health interventions in those provinces. The research was evident that breast, prostate, and lung cancers showed the highest incidences, meanwhile colorectal and urinary system cancer types had comparatively lower rate. Our findings underscore the significance of ameliorating environmental and demographic factors which could contribute to these patterns. Future studies however could focus on identifying specific environmental or health factors driving these spatial disparities to guide policy and prevention efforts effectively.

Declarations

Conflicts of Interest

The author declares no conflicts of interest.

Funding

This research received no external funding.

Data Availability Statements

The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

[1]. Quinn, R. M., Bernal, A. M., Oh, S. Y., & Anampa, J. D. (2024). Trends in incidence of invasive lobular carcinoma of the breast by race: Patterns by age, cancer stage, and socioeconomic factors in the United States, 1992-2019. *Clinical Breast Cancer*. https://doi.org/10.1016/j.clbc.2024.12.015

[2]. Gupta, A., & Akinyemiju, T. (2024). Early-onset cancer incidence in the United States by race/ethnicity between 2011 and 2020. *Cancer Epidemiology*, *92*. https://doi.org/10.1016/j.canep.2024.10263

[3]. Kava, C. M., Siegel, D. A., Sabatino, S. A., Qin, J., Richards, T. B., & Jane Henley, S. (2024). Lung cancer incidence, 2019–2020, United States: The potential impact of the COVID-19 pandemic. *Annals of Epidemiology*, *98*, 44–50. https://doi.org/10.1016/j.annepidem.2024.08.005

[4]. Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, *73*(1), 17–48. https://doi.org/10.3322/caac.21763

[5]. Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., … Chen, W. (2022). Cancer statistics in China and United States, 2022: Profiles, trends, and determinants. *Chinese Medical Journal*, *135*(5), 584–590. https://doi.org/10.1097/CM9.0000000000002108

[6]. Ma, F., Lehnherr, M., Fornoff, J., & Shen, T. (2011). Childhood cancer incidence in proximity to nuclear power plants in Illinois. *Archives of Environmental and Occupational Health*, *66*(2), 87–94. https://doi.org/10.1080/19338244.2010.506497

[7]. Keller, J. E., & Howe, H. L. (1994). *Case-Control Studies of Cancer in Illinois Farmers Using Data From the Illinois State Cancer Registry and the U.S. Census of Agriculture*. *EurJ Cancer* (Vol. 30).

[8]. Guo, L., Wright, M. E., Osias, M. C., Vaezi, M., & Hughes, M. C. (2022). Creation and Evaluation of the Illinois Cancer Risk Index as a Predictor of Four Common Cancers. *Preventing Chronic Disease*, *19*. https://doi.org/10.5888/PCD19.220104

[9]. Guo, L., Wright, M. E., Osias, M. C., Vaezi, M., & Hughes, M. C. (2022). Creation and Evaluation of the Illinois Cancer Risk Index as a Predictor of Four Common Cancers. *Preventing Chronic Disease*, *19*. https://doi.org/10.5888/PCD19.220104

[10]. Singh, G. K., & Jemal, A. (2017). Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950-2014: Over Six Decades of Changing Patterns and Widening Inequalities. *Journal of Environmental and Public Health*, *2017*. https://doi.org/10.1155/2017/2819372

[11]. Yang, J., Lou, S., & Yao, T. (2024). Trends in primary malignant bone cancer incidence and mortality in the United States, 2000–2017: A population-based study. *Journal of Bone Oncology*, *46*. https://doi.org/10.1016/j.jbo.2024.100607

[12]. Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., … Chen, W. (2022). Cancer statistics in China and United States, 2022: Profiles, trends, and determinants. *Chinese Medical Journal*, *135*(5), 584–590. https://doi.org/10.1097/CM9.0000000000002108

[13]. Qiu, H., Cao, S., & Xu, R. (2021). Cancer incidence, mortality, and burden in China: a time-trend analysis and comparison with the United States and United Kingdom based on the global epidemiological data released in 2020. *Cancer Communications*, *41*(10), 1037–1048. https://doi.org/10.1002/cac2.12197

[14]. Wang, F., Guo, D., & McLafferty, S. (2012). Constructing geographic areas for cancer data analysis: A case study on late-stage breast cancer risk in Illinois. *Applied Geography*, *35*(1–2), 1–11. https://doi.org/10.1016/j.apgeog.2012.04.005