Global Scientific JOURNALS

GSJ: Volume 11, Issue 9, September 2023, Online: ISSN 2320-9186
www.globalscientificjournal.com

Incidence and Mortality Trend Analysis using Historical Cancer Data of the United States of America

Mojeed Dayo Ajegbile[1], Emmanuel Kanu[2], Yusuff Bolaji Ajegbile[3], Joy Adedeji[4], Cecilia Dada[5], Olowoopejo Olanrewaju[6], Miracle Agboluaje[7], Funmilayo Olaseni[8]

[1]Department of Computer Science and Quantitative Methods Austin Peay State University, Clarksville, TN, USA
[2]Department of Computer Science and Quantitative Methods Austin Peay State University, Clarksville, TN, USA
[3]Department of Health Education University of Ibadan, Nigeria
[4] Department of Information Technology, University of the Cumberlands, KY, USA
[5]Department of Zoology, University of Ibadan, Nigeria
[6]Department of Zoology, University of Lagos, Nigeria
[7]Department of Healthcare Administration, California State University, Bakersfield
[8]Department of Zoology, Federal University of Agriculture, Abeokuta, Nigeria

Abstract

Cancer is a complex disease that affects people of different ages, ethnicities, sexes, and races differently. To address the disparities in cancer outcomes among different populations, there is a need for a better understanding of the factors that contribute to these disparities. This study analyzed data from the CDC's National Program of Cancer Registries (NPCR) from 1999 to 2014 to examine how cancer affects different races and sexes. Our dataset included 961,776 cancer cases across 14 columns.
Our findings suggest that there are significant differences in cancer incidence and mortality rates among different racial and ethnic groups. For instance, African Americans have higher cancer incidence and mortality rates than any other racial or ethnic group in the US. Additionally, we found that there are differences in cancer incidence and mortality rates among different sexes. For example, men have higher incidence rates of cancer than women, but women have a slightly higher mortality rate from cancer. Our study underscores the need for targeted interventions to reduce the burden of cancer in high-risk populations. By identifying the factors that contribute to disparities in cancer outcomes, we can develop effective strategies to reduce cancer morbidity and mortality in all populations.

The aim of the cancer analysis study was to examine comparatively the cancer trends in the United States between 1999 and 2014. To do this, we gathered cancer registry data from two primary sources, the Center for Disease Control and Protection (CDC) and the National Cancer Institute (NCI), as well as cancer mortality data from the CDC's National Center for Health Statistics. These sources provided comprehensive information on new cancer cases and cancer deaths for the entire U.S. population.
Since we are interested in national trends rather than regional ones, we filtered out any data that was specific to regions. This subset of data allowed the research to focus analysis on cancer trends at a national level, and to identify any patterns or changes in cancer rates across the United States during the study period. Gathering data from reliable sources is a crucial component of any research study and is especially important in cancer research. The CDC and NCI are highly respected organizations in the field of cancer research, and their cancer registry data is widely used in studies investigating cancer trends and outcomes. Similarly, the CDC's National Center for Health Statistics is a reputable source of mortality data in the United States. By using data from these sources, we can draw on high-quality, reliable information to inform the analysis of cancer trends in the United States.

Introduction

Cancer is a major public health issue in the United States, with billions of dollars being spent on prevention, prophylactic, and therapeutic measures to control the disease. Despite these efforts, cancer continues to affect people of different ages, ethnicities, sexes, and races, but not equally. Genetic, hormonal, and environmental factors can lead to differences in cancer risk among ethnic groups. Understanding these differences is crucial for developing targeted interventions to reduce cancer incidence and mortality. In this paper, we aim to project how cancer affects different races and sexes using data from the Centre for Disease Control and Prevention (CDC) from 1999 to 2014.

According to [1], the most diagnosed cancers in the United States include breast, lung, prostate, colorectal, and melanoma. These cancers account for a substantial portion of cancer cases in the country. There are variations and disparities in cancer mortality rates across different demographic groups and geographical regions in the United States. Factors such as socioeconomic status, access to healthcare, and disparities in cancer screening and treatment contribute to these variations [2].

The incidence rates of certain cancers have been changing over time. For example, lung cancer incidence has been decreasing in recent years due to a decline in smoking rates, while the incidence of some cancers, such as liver and thyroid cancer, has been increasing [3]. However, Black Americans have a higher incidence of cancer than White Americans for most cancer types, including breast, lung, and colorectal cancer. Hispanic

Americans also have a higher incidence of cancer than White Americans for some cancer types, such as cervical cancer [4, 17].

Overall cancer mortality rates in the United States have been declining over the past few decades. This decline can be attributed to advancements in cancer treatment, increased screening efforts, and reductions in tobacco use [5]. Black Americans have a higher mortality rate from cancer than White Americans for most cancer types. Hispanic Americans also have a higher mortality rate from cancer than White Americans for some cancer types, such as cervical cancer [6, 16].

Racial and Ethnic Disparities in Cancer Mortality have also been well-researched. For instance, African Americans have higher lung cancer mortality rates compared to other racial and ethnic groups, partly due to a higher prevalence of smoking and disparities in access to healthcare and early detection [7]. African American men have higher prostate cancer incidence and mortality rates compared to white men. These disparities may be influenced by genetic factors, socioeconomic factors, and differences in healthcare access [8].

Racial and ethnic disparities in cancer incidence and mortality in the United States highlight the need for targeted interventions to reduce these inequities. Efforts should focus on improving access to healthcare services, promoting cancer screening and early detection, and addressing socioeconomic and cultural factors that contribute to disparities. By implementing comprehensive strategies, we can work towards achieving health equity in cancer care for all populations.

Literature Review

Cancer is a major public health concern in the United States, with significant implications for morbidity and mortality. Understanding the incidence and mortality rates, as well as the trends associated with different cancer types, is crucial for effective cancer control strategies and healthcare resource allocation. This literature review aims to consolidate existing knowledge on cancer incidence, mortality, and trends in the United States, shedding light on the current state of the cancer burden.

Cancer Incidence:
Several population-based cancer registries provide valuable data on cancer incidence rates. The Surveillance, Epidemiology, and End Results (SEER) program conducted by the National Cancer Institute (NCI) is one of the primary sources of cancer incidence data in the United States. Analysis of SEER data reveals variations in cancer incidence rates across different regions and demographic groups. Studies have shown increased incidence rates for certain cancers, such as breast, lung, prostate, and colorectal cancer.

Environmental factors, lifestyle choices, and genetic predisposition contribute to these variations [9].

Cancer Mortality:
Examining cancer mortality rates provides insights into the impact of cancer on the population's health. The National Vital Statistics System and the Centers for Disease Control and Prevention (CDC) serve as essential sources for mortality data. Over the years, advancements in early detection and treatment have contributed to a decline in overall cancer mortality rates. However, disparities persist among different cancer types and demographic groups. Lung cancer remains the leading cause of cancer-related deaths, followed by colorectal, breast, and prostate cancers. Disparities in mortality rates can be attributed to factors such as late-stage diagnosis, limited access to quality healthcare, and variations in treatment effectiveness [10].

Cancer Trends:
Analyzing cancer trends is crucial for identifying shifts in the cancer burden and evaluating the impact of prevention and control measures. The examination of long-term trends has revealed significant changes in the incidence and mortality rates of specific cancers. For instance, declines in lung cancer incidence and mortality rates are observed due to anti-smoking campaigns and improved treatment options. Conversely, certain cancers, such as liver and melanoma, have shown increasing incidence rates, emphasizing the need for targeted prevention and early detection strategies [11].
This literature review provides an overview of cancer incidence, mortality, and trends in the United States. The synthesis of available data highlights the significance of ongoing efforts in cancer prevention, early detection, and treatment. Understanding the evolving landscape of cancer burden is crucial for implementing effective interventions, reducing disparities, and improving outcomes for individuals affected by cancer.

Method

Data Aggregation and Cleaning:
In this study, Python programming language was employed for data aggregation, data cleaning, and subsequent analysis. The data used in this research was obtained from a comprehensive database consisting of cancer patient records. The Python libraries, Numpy and Scipy, were utilized for efficient data manipulation and statistical calculations. The data aggregation process involved extracting relevant information such as patient demographics, cancer type, and clinical outcomes.

Data Analysis:
After the data aggregation and cleaning steps, the dataset was subjected to exploratory data analysis to gain insights into the variables and identify any patterns or trends.

Descriptive statistics, such as mean, median, standard deviation, and frequency distributions, were computed using Numpy and Scipy. These statistical analyses provided a comprehensive overview of the dataset, allowing for a better understanding of the characteristics of the cancer cases under investigation.

Logistic Regression Model:
To predict the possibility of a specific cancer type leading to mortality over an extended period, a logistic regression model was employed. This predictive model, implemented using Scipy, aimed to determine whether a given cancer type would cause death or not. The logistic regression algorithm is well-suited for binary classification tasks, such as predicting mortality outcomes. The input variables for the model included cancer type, patient age, gender, and other relevant clinical parameters available in the dataset.

Model Training and Evaluation:
The logistic regression model was trained using a subset of the dataset, with appropriate features and corresponding mortality outcomes. The dataset was randomly split into training and testing sets, ensuring that the model was trained on a sufficiently diverse range of cases. The training set was used to fit the logistic regression model, optimizing the model parameters to minimize the prediction errors. The performance of the trained model was then evaluated using the testing set, employing relevant evaluation metrics such as accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC).

Statistical Significance

To assess the statistical significance of the logistic regression model and the predictive power of the included variables, appropriate statistical tests, such as chi-square tests or t-tests, were conducted using Scipy. These tests determined the significance of the relationships between the input variables and the mortality outcomes, thus providing insights into the predictive capabilities of the logistic regression model.
In summary, this study utilized Python programming language, along with Numpy and Scipy libraries, for data aggregation, data cleaning, and subsequent analysis. A logistic regression model was implemented using Scipy to predict the possibility of a cancer type leading to mortality over a long period. The model was trained, evaluated, and statistically analyzed to assess its predictive capabilities and the significance of the included variables.

Several questions related to cancer incidence, trends, and mortality are intended to be answered through this analysis. Some of the most germane questions are:

(1) What are the most prevalent cancer types in the United States of America?

In the United States of America, determining the most prevalent cancer types is of great importance for public health planning and resource allocation. Understanding the

specific types of cancer that affect the population at a higher frequency can aid in targeting prevention strategies, improving early detection efforts, and developing effective treatment protocols. By identifying the most prevalent cancer types, healthcare systems can allocate resources, research funding, and public awareness campaigns more efficiently.

(2) What are the deadliest cancer types across races and gender?

Understanding the cancer mortality rate in the United States is crucial for assessing the impact of cancer on the population's health and guiding public health interventions. The cancer mortality rate represents the number of deaths attributed to cancer per unit of the population over a specific period. By quantifying the burden of cancer-related deaths, healthcare systems can allocate resources, implement preventive measures, and improve treatment strategies to reduce mortality and improve overall survival rates.
To determine the cancer mortality rate in the United States, comprehensive data from national vital statistics systems, cancer registries, and population-based studies are utilized. These sources provide information on the number of cancer-related deaths, as well as demographic characteristics such as age, gender, and race/ethnicity. The data are often analyzed in conjunction with population estimates to calculate mortality rates per 100,000 individuals.

(3) What is the cancer distribution by race in the United States of America?

Understanding the distribution of cancer by race in the United States of America is essential for identifying disparities, addressing healthcare inequities, and tailoring interventions to specific populations. Cancer incidence and mortality rates can vary among different racial and ethnic groups, highlighting the importance of analyzing the burden of cancer on a racial basis.
To assess the distribution of cancer by race in the United States, data from various sources, including cancer registries, national surveys, and research studies, are analyzed. These sources provide information on cancer diagnoses and deaths, as well as demographic characteristics such as race and ethnicity. The data are often stratified by race to determine the specific cancer burden within each racial group.

(4) What is the cancer incidence and mortality trend in the United States?

Understanding the cancer incidence and mortality trends in the United States is essential for assessing the overall burden of cancer, tracking progress in cancer prevention and treatment, and identifying areas for targeted interventions. Examining these trends over time provides valuable insights into the changing patterns of cancer occurrence and the effectiveness of public health efforts.
Cancer incidence refers to the number of new cases of cancer diagnosed within a given population over a specific period. Monitoring cancer incidence trends helps identify shifts in the prevalence of different cancer types and can indicate changes in risk factors,

screening practices, and environmental exposures. What's the possibility of a cancer type resulting in death?

The possibility of a cancer type leading to death depends on various factors, including the specific cancer type, stage of cancer, treatment options, and individual patient characteristics. Cancer encompasses a diverse group of diseases, each with its own prognosis and potential for mortality.

Some cancer types have higher mortality rates due to their aggressive nature, limited treatment options, or late-stage diagnosis. For example, cancers such as pancreatic, liver, and lung cancer often have lower survival rates and a higher likelihood of leading to death. These cancers tend to be more challenging to detect early and may exhibit rapid progression.

On the other hand, certain cancer types, especially those detected at an early stage, can have better prognoses and higher chances of long-term survival. Cancers like breast, prostate, and skin cancer, when detected early and treated appropriately, often have higher survival rates and a lower likelihood of leading to death.

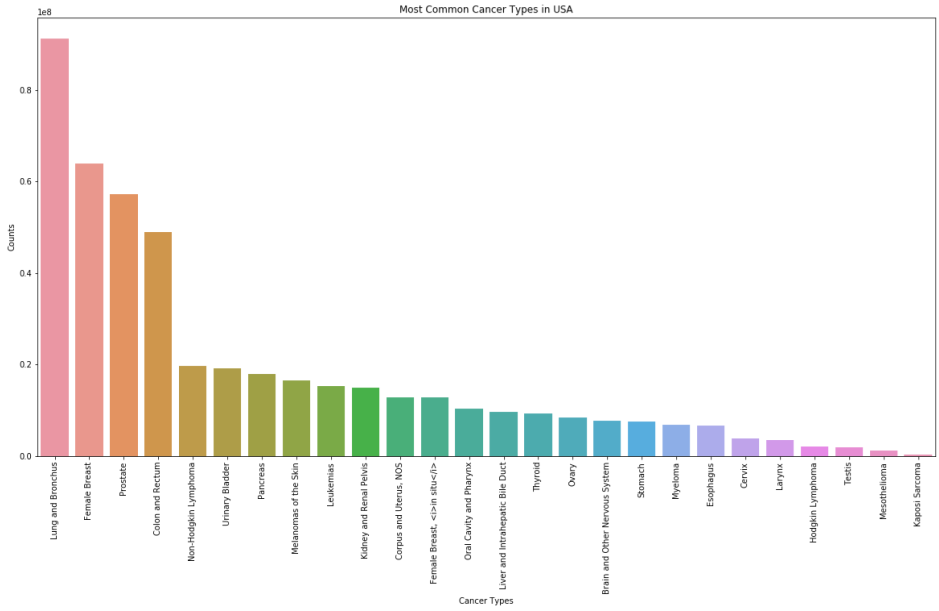
Advancements in cancer research and treatment options have contributed to improved outcomes and increased survival rates for many cancer types. Personalized medicine approaches, targeted therapies, immunotherapy, and advances in surgical techniques have helped extend life expectancy and improve the quality of life for patients with cancer.

It is important to note that survival rates and the possibility of death can also vary within cancer types based on individual patient characteristics, such as age, overall health, genetic factors, and response to treatment. Therefore, it is essential for individuals diagnosed with cancer to work closely with their healthcare team to understand their specific prognosis, treatment options, and potential outcomes.

While statistical data and survival rates provide general insights, they should be interpreted with caution and personalized medical advice sought for a comprehensive understanding of the possibility of a specific cancer type leading to death in an individual case.

Result

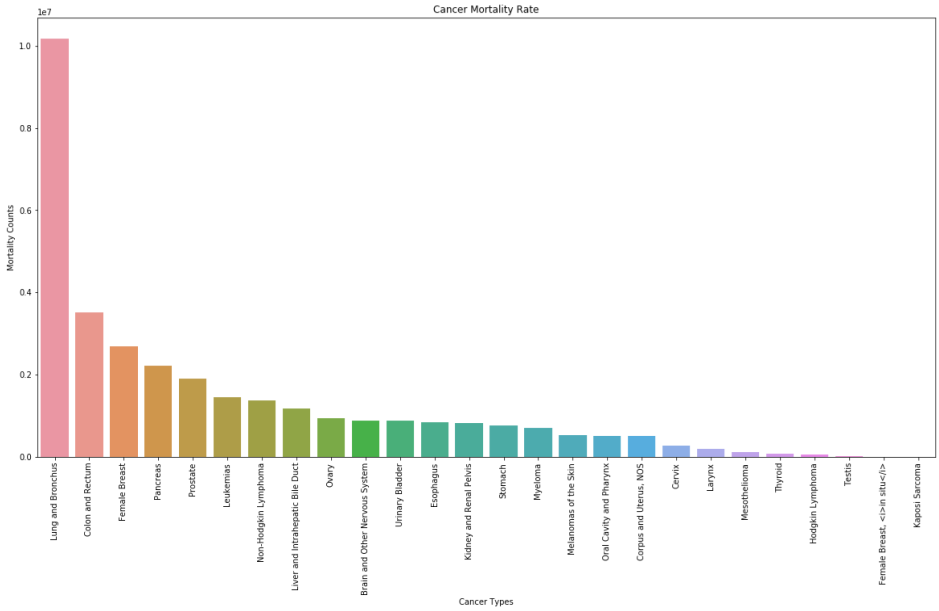
Our analysis showed that Lung and Bronchus, Female Breast and Prostate Cancer types are the most common in the United States, while Testis, Mesothelioma, and Kaposi Sarcoma are the least prevalent among the 26 types of cancer analyzed. However, recent studies, suggest that with new cases rising in the United States in 2023, breast cancer, and prostate cancer are leading lung cancer due to higher incidences of these two cancer types. Also, several research into the lifesaving potentials of lung cancer screening, which can detect the disease at an earlier stage when it's more curable, and advanced lung cancer research might have contributed to a decline in lung cancer incidence rate [13]. It is recorded that the rate of new lung cancer cases has decreased by 11% nationally [19].



Most prevalent Cancer Types from 2009 to 2014

Cancer Mortality from 2009 to 2014

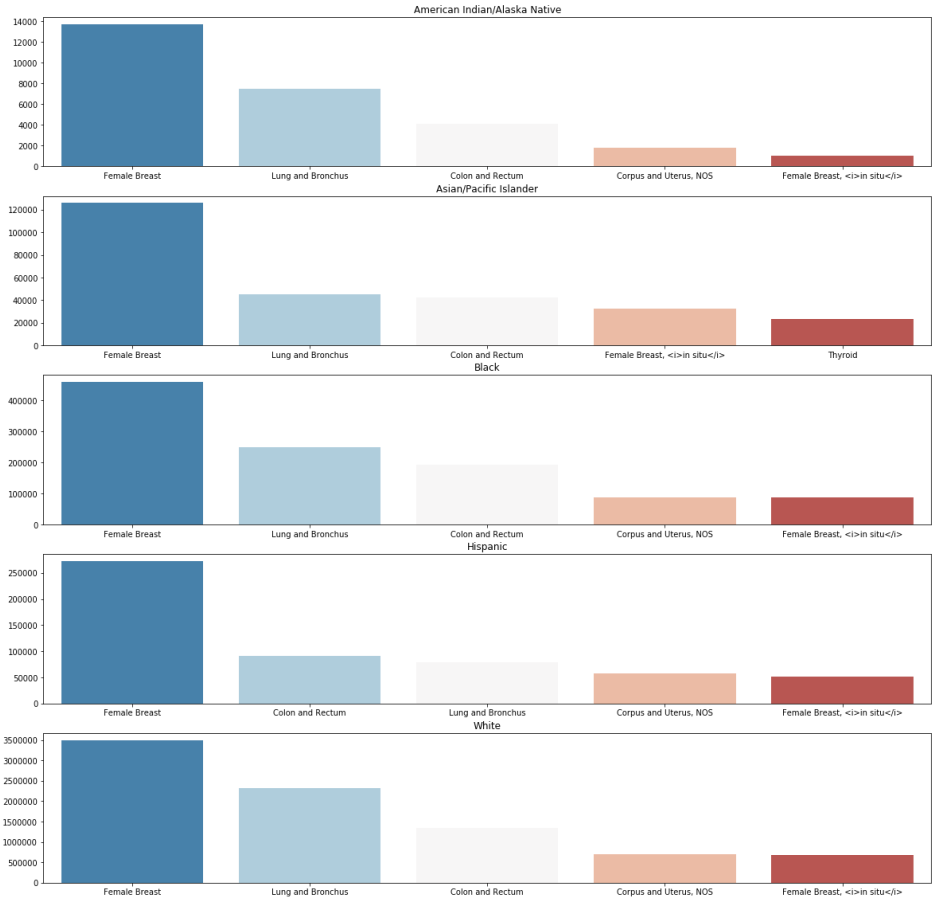
Our analysis shows Lung and Bronchus cancer killed the most within the years examined. This is followed by Colon and Rectum and Female Breast Cancers. This analysis supports the report published by the Centre for Disease Prevention and Control (CDC), which shows that in the year 2020, lung cancer is still the cancer type that has the highest mortality rates, followed by colon and rectum, pancreas, and female breast cancer. [18]



Cancer Mortality Rate Between 2009 to 2014

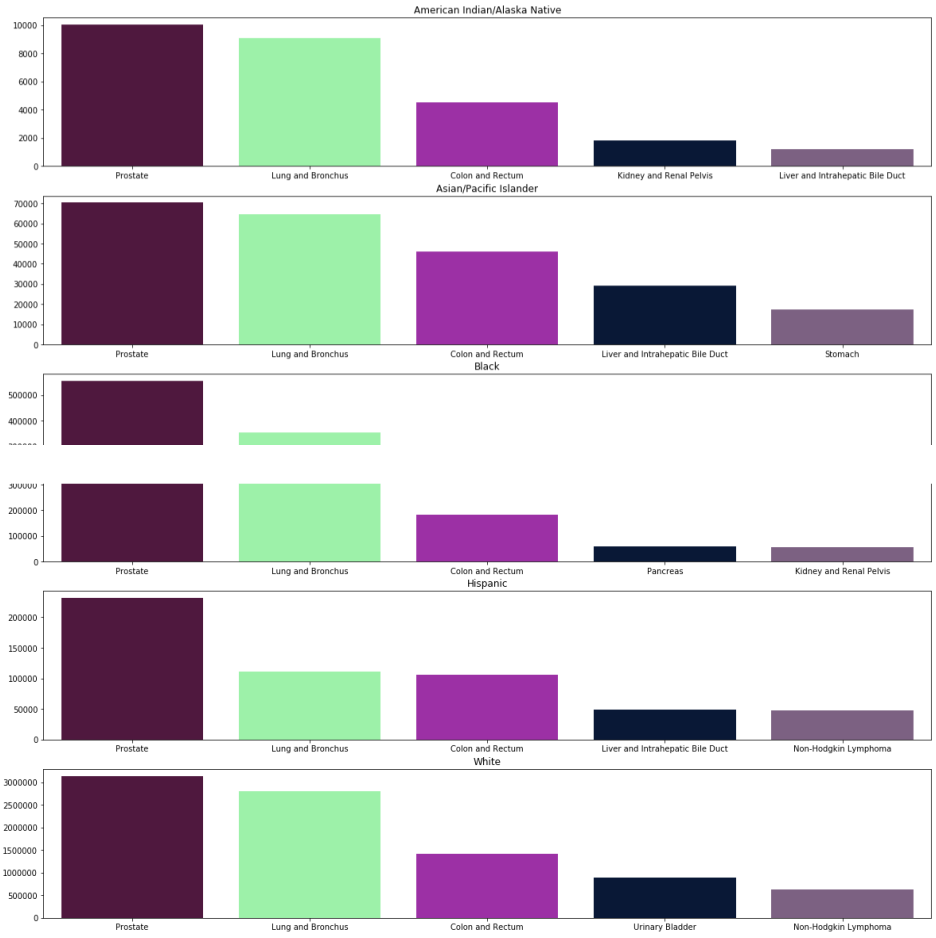
Cancer Distribution Across Sex

Our analysis suggests that irrespective of Race, Female breast cancer is the leading cause of female cancer incidence. This is followed by Lung and Bronchus Cancer. This suggests that irrespective of race, color, or ethnicity, Breast Cancer is of serious public health concern among females.



Cancer Distribution Across Female Gender by Race.

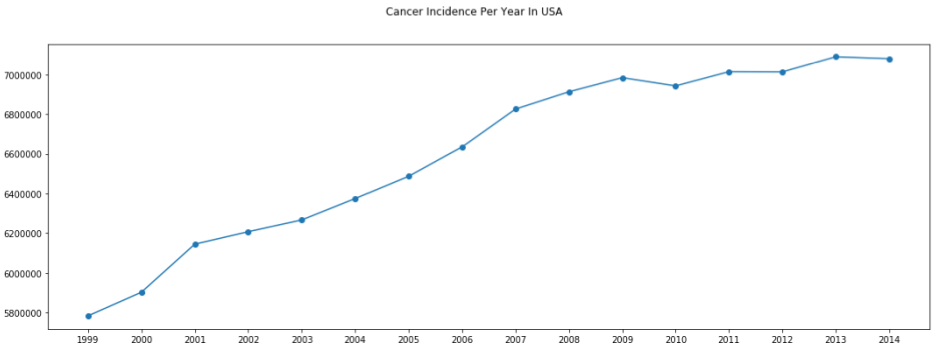
Among males of any race and ethnicity, Prostate, Lung and Bronchus cancer are the types that have the highest mortality rates.



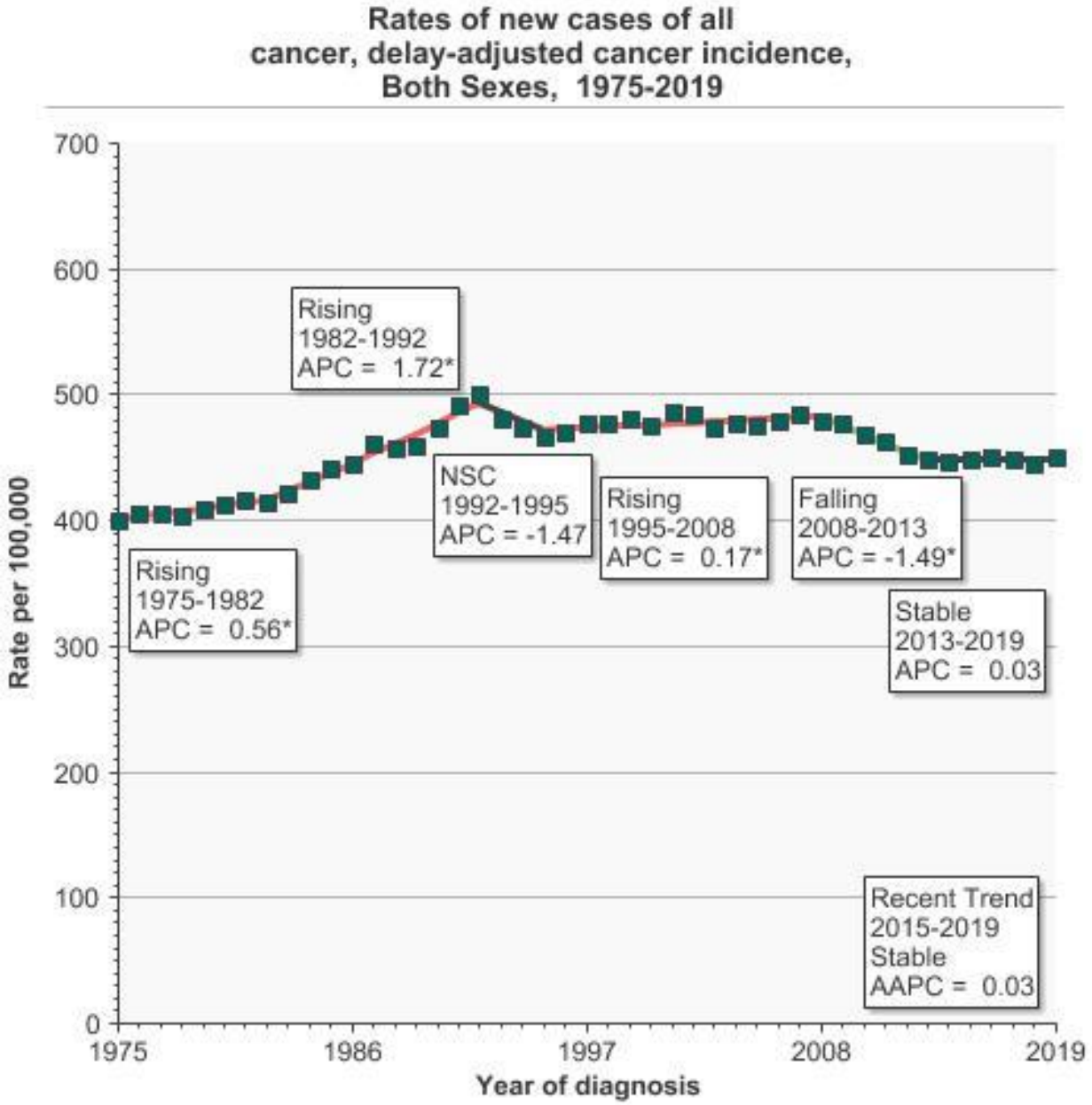
Cancer Distribution in Males by Race

Cancer Trend

Our study revealed that from 2010 to 2014, the cumulative cancer incidence trend in the United States decreased slightly. This agrees with the current National Cancer Institute report suggesting a stable cancer incidence rate.



Cancer cumulative Incidence Trend in the United States between 2009 to 2014.

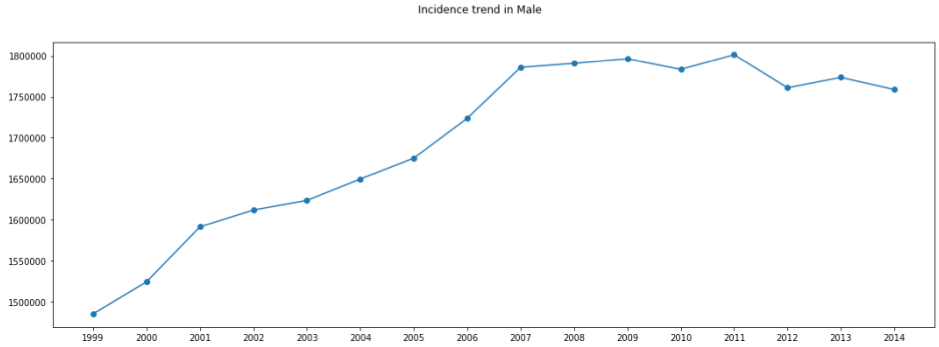


No Healthy People Target for this measure.
Source: SEER Program, National Cancer Institute. Incidence data are from the SEER 8 areas (http://seer.cancer.gov/registries/terms.html).
Data are age-adjusted to the 2000 US standard population using age groups: <1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+.
Weighted regression lines are calculated using the Joinpoint Trend Analysis Software, Version 4.9 April 2022, National Cancer Institute.
* The Annual Percent Change (APC)/Average Annual Percent Change (AAPC) is statistically significant.
The AAPC is the Average Annual Percent Change and is based on the APCs calculated by Joinpoint.
NSC: Non-Significant Change.

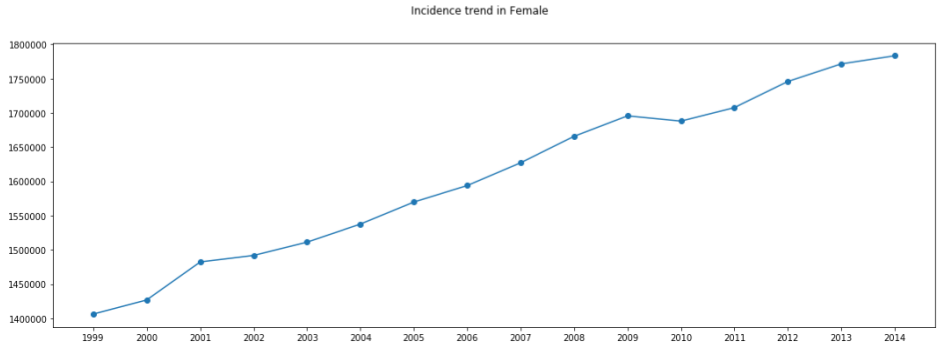
National Cancer Institute "Cancer Incidence Rates for Both Sexes"

Our analysis also suggested that while new cancer cases are reduced in males, they might be increasing in females. Our cancer incidence trend suggests a sharp decline in cancer incidence

but an increase in the female incidence of new cases. The cause of this will be further investigated in future studies.

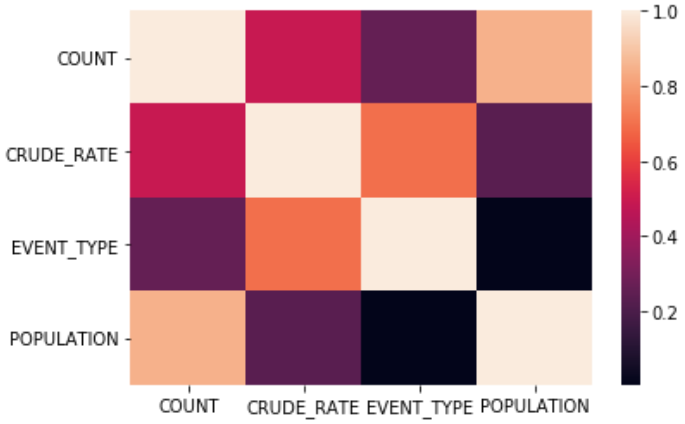


Cancer Incidence in Males

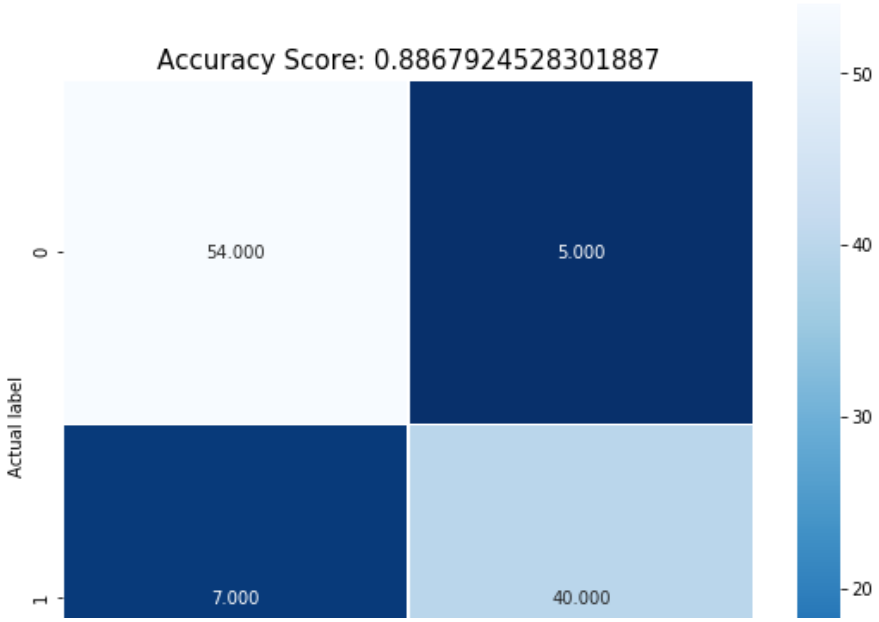


Cancer Incidence in Females.

Our analysis shows that 88% of the cancer types that were considered could potentially lead to death if not properly treated.



Features Considered for the logistic regression problem.



Confusion Matrix for Classifying Possible Outcome of a Cancer Type

Discussion

The prevalence of cancer types can vary over time and across different demographic groups. Factors such as age, gender, race/ethnicity, and geographic location can influence the distribution of cancer types within a population. Therefore, comprehensive and up-to-date data sources are crucial for accurate assessments. Surveillance systems, cancer registries, and large-scale studies play a significant role in providing valuable information on cancer prevalence trends.
To determine the most prevalent cancer types in the United States, data from various sources are analyzed, including national cancer registries, population-based surveys, and research studies. These sources provide detailed information on the incidence and prevalence of different cancer types across diverse populations. Statistical methods, such as prevalence rates and standardized rates, are employed to account for variations in population size and demographics.

Cancer mortality rates can vary by cancer type, age group, gender, and geographic region. For instance, certain types of cancer, such as lung and colorectal cancer, have higher mortality rates due to their aggressive nature and challenges in early detection. Additionally, disparities in cancer mortality rates may exist among different racial and ethnic groups, with certain populations experiencing higher mortality rates compared to others.
Monitoring the cancer mortality rate over time allows for identifying trends and evaluating the effectiveness of cancer control measures. These measures may include advancements in screening programs, improvements in treatment options, and increased public awareness of risk factors and early detection methods.

The distribution of cancer by race in the United States can reveal variations in the prevalence of specific cancer types among different racial and ethnic populations. For instance, certain cancers, such as prostate cancer, are more prevalent among African American men, while breast cancer has higher incidence rates among white women. Additionally, disparities may exist in cancer outcomes, with some racial and ethnic groups experiencing higher mortality rates compared to others.
Analyzing the distribution of cancer by race can also shed light on the influence of socioeconomic factors, access to healthcare, and cultural differences on cancer incidence and outcomes. It provides insights into disparities in risk factors, screening rates, treatment options, and overall quality of care.
Efforts to address racial disparities in cancer distribution and outcomes often involve targeted interventions. These may include increasing awareness and access to cancer screenings, promoting culturally sensitive healthcare services, improving health education, and addressing socioeconomic barriers that contribute to disparities.

Some cancer types may show increasing incidence rates due to factors such as population aging, changes in lifestyle behaviors, or improved diagnostic techniques. Conversely, declining incidence rates may be observed for certain cancers due to successful prevention strategies or changes in risk factor prevalence.
Cancer mortality, on the other hand, represents the number of deaths caused by cancer within a population. Examining trends in cancer mortality rates provides insights into improvements in early detection, treatment advancements, and changes in risk factor management. Decreasing mortality rates indicate progress in cancer care, including better survival outcomes and improved access to screening and treatment services. Conversely, stable or increasing mortality rates may highlight challenges in addressing specific cancers or disparities in healthcare access.
Understanding the incidence and mortality trends for specific cancer types is crucial, as different cancers may exhibit unique patterns. For example, lung cancer incidence and mortality rates have generally been declining in recent years due to reduced smoking rates and improved treatment options. However, certain cancers like liver cancer and melanoma have shown increasing incidence rates, underscoring the need for targeted prevention and early detection efforts.
Monitoring cancer incidence and mortality trends allows policymakers, researchers, and healthcare professionals to assess the effectiveness of cancer control strategies, allocate resources appropriately, and tailor interventions to address the evolving cancer landscape. It helps inform public health policies, screening guidelines, and treatment approaches to further reduce the burden of cancer and improve outcomes for individuals and communities across the United States.

References

1. National Cancer Institute (NCI). Surveillance, Epidemiology, and End Results (SEER) Program. Available at: https://seer.cancer.gov/statfacts/html/common.html
2. National Cancer Institute. About Cancer Health Disparities. Available at https://www.cancer.gov/about-nci/organization/crchd/about-health-disparities
3. National Cancer Institute. Cancer Statistics. Available at: https://www.cancer.gov/about-cancer/understanding/statistics
4. Centers for Disease Control and Prevention. African American People and Cancer. Available at: https://www.cdc.gov/cancer/health-equity/groups/african-american.htm
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2021. CA: A Cancer Journal for Clinicians. 2021;71(1):7-33.
6. American Cancer Society. Cancer Disparities in the Black Community. Available at: https://www.cancer.org/about-us/what-we-do/health-equity/cancer-disparities-in-the-black-community.html
7. Brid M. Ryan, Lung cancer health disparities, Carcinogenesis, 2018, Vol. 39, No 6, 741-751
8. Dallin Lowder et. al, Racial disparities in prostate cancer: A complex interplay between socioeconomic inequities and genomics, Cancer Letters 531 (2022) 71-82
9. Cancer Disparities. National Cancer Institute. Available at https://www.cancer.gov/about-cancer/understanding/disparities
10. Wayne J. Riley, Health Disparities: Gaps in Access, Quality and Affordability of Medical Care, Trans Am Clin Climatol Assoc. 2012; 123: 167–174.
11. Erin P. Balogh et al, Reducing Tobacco-Related Cancer Incidence and Mortality: Summary of an Institute of Medicine Workshop, Oncologist. 2014 Jan; 19(1): 21–31.
12. American Cancer Society. Cancer Facts & Figures 2021. Available at: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2022.html
13. Howlader N, et al.The Efect of Advances in Lung-Cancer Treatment on Population Mortality. N Engl J Med 2020; 383:640-649
14. Brid M. Ryan, Lung cancer health disparities, Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Betheda, MD 20892, USA
15. Eileen B. O'Keefe et al. Health Disparities and Cancer: Racial. Disparities in Cancer Mortality in the United States, 2000-2010, Frontier in Public Health
16. Valentina A. Zavala et al, Cancer health disparities in racial/ethnic minorities in the United States, Br J Cancer. 2021 Jan 19; 124920: 315-332
17. Angela N. Giaquinto et al, Cancer statistics for African American/Black People 2022, CA: A cancer Journal for Clinicians/Volume 72, Issue 3/ p. 202-229
18. National Cancer Institute- Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: common Cancer Sites. Available at: https://seer.cancer.gov/statfacts/html/common.html#:~:text=Lung%20and%20bronchus%20cancer%20is%20responsible%20for%20the%20most%20deaths,deadliest%20cancer%2C%20causing%2050%2C550%20deaths

19. American Cancer Society. Key Statistics for Lung Cancer. Available at https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html