U.S. Cancer Statistics

Data Visualizations Tool Technical Notes

2023 Submission
Diagnosis Years 1999–2021
Released June 2024

Contents

Introduction

The impact of cancer

Cancer is the second-leading cause of death in the United States[1] and 608,366 people living in the United States died from this disease in 2022.[2] The 2024 release of United States Cancer Statistics data indicates that, in 2021 (the most recent year of incidence data available), 1,777,566 people living in the United States received a new diagnosis of invasive cancer. These counts do not include *in situ* cancers, benign and borderline brain and central nervous system tumors, and basal and squamous cell skin cancers.[A] As of January 1, 2021, an estimated 14 million people living in the United States were alive with a history of invasive cancer diagnosed in the past 20 years.[3]

The Agency for Healthcare Research and Quality's (AHRQ) Medical Expenditure Panel Survey estimates that for 2021, the direct medical costs for cancer, including all health care expenditures, were $146 billion.[3]

Cancer prevention

Several primary and secondary prevention measures could substantially reduce the number of new cancer cases and prevent many cancer-related deaths. To reduce the nation's cancer burden, the Centers for Disease Control and Prevention's (CDC's) Division of Cancer Prevention and Control (DCPC) aims to monitor and reduce factors that may increase cancer risk. CDC supports partners to make high-quality screening services and referral to treatment available and accessible to everyone, including medically underserved populations.[4] [5] CDC supports the development of comprehensive cancer control plans,[6] which include proven strategies and planned actions to prevent cancer.

How cancer data are collected

Incidence data

Cancer is a mandatory reportable disease in the United States. Cancer registries collect population-based data about:[7] [8]

- The occurrence of cancer (incidence).
- The types of cancer (morphology).
- The site in the body where the cancer first occurred (primary site).
- The extent of disease at the time of diagnosis (stage).
- The planned first course of treatment.
- [1]The outcome of treatment and clinical management (survival and vital status).

These cancer reports are sent to central cancer registries at the state or territory level from a variety of medical facilities, including hospitals, physicians' offices, radiation facilities, freestanding surgical centers, and pathology laboratories.

[A]Data are from selected central cancer registries that meet the data quality criteria for all reportable cancer sites combined. These data cover 98% of the U.S. population for 2021 diagnoses and 99.6% for cases diagnosed between 1999 and 2021. See registry-specific data quality information.

Death (mortality) data

Death data, including deaths due to cancer, are recorded on death certificates that are sent to state or territory vital statistics offices. Death data include information regarding primary cancer site, and may also include morphology, according to International Classification of Diseases, Tenth Revision (ICD-10).

Uses of cancer data

Cancer data are critical for directing effective cancer prevention and control programs. These programs focus, in part, on preventing behaviors that may put people at an increased risk for cancer (such as tobacco use), and on reducing environmental risk factors (such as exposure to things known to cause cancer).

Cancer information is essential for deciding where to have cancer screening programs and making long-term plans for diagnostic and treatment services. Data at the national, state or territory, congressional district, and county levels help public health officials prioritize and monitor public health efforts and track progress toward Healthy People objectives.

Contributors

CDC's National Program of Cancer Registries (NPCR), NCI's Surveillance, Epidemiology, and End Results (SEER) Program, and CDC's National Vital Statistics System (NVSS) contributed to U.S. Cancer Statistics data.

Partners crucial to the success of cancer registration and cancer surveillance in the United States include the American Cancer Society, the American College of Surgeons, the American Joint Committee on Cancer, the National Cancer Registrars Association, and the North American Association of Central Cancer Registries.

References

1. Kochanek KD, Murphy SL, Xu JQ, Arias E. Mortality in the United States, 2022. NCHS Data Brief, no 492. Hyattsville, MD: National Center for Health Statistics. March 2024.
2. U.S. Cancer Statistics Data Visualizations Tool. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, released in June 2024.
3. Agency for Healthcare Research and Quality. Total expenditure ($) in millions by condition, United States, 2021. Medical Expenditure Panel Survey. Generated interactively: May 19, 2024.
4. Curry SJ, Byers T, Hewitt M. *Fulfilling the Potential of Cancer Prevention and Control*. Washington (DC): The National Academies Press; 2003.
5. Haynes MA, Smedley BD. *The Unequal Burden of Cancer: An Assessment of NIH Research and Programs for Ethnic Minorities and the Medically Underserved*. Washington (DC): The National Academies Press; 1999.
6. National Comprehensive Cancer Control Program. About comprehensive cancer control programs. Atlanta: U.S. Department of Health and Humans Services, Centers for Disease Control and Prevention. Accessed May 16, 2024.

7. American College of Surgeons Commission on Cancer. Standards for Oncology Registry Entry (STORE). Accessed May 1, 2024.

8. National Cancer Institute. SEER Program Coding and Staging Manual 2023 and 2024. Accessed May 1, 2024.

Contributors

National Program of Cancer Registries (NPCR)

CDC's National Program of Cancer Registries (NPCR) funds 50 cancer registries: 46 states, the District of Columbia, Puerto Rico, the Pacific Island Jurisdictions, and the U.S. Virgin Islands.

NPCR is committed to:

- Monitoring the state and national burden of cancer.
- Identifying variation in cancer incidence for racial and ethnic populations and for regions within a state, between states, and between regions.
- Providing data for research.
- Providing guidance for the allocation of health resources.
- Responding to public concerns and inquiries about cancer.
- Improved planning for future health care needs.
- Evaluating activities in cancer prevention and control.

In January 2001, NPCR-funded registries began reporting their incidence data annually to CDC. The registries report data to CDC beginning with cases diagnosed in the first year for which they collected data with the assistance of NPCR funds. Data from the special population cancer registries or the SEER metropolitan area cancer registries operating in Alaska, Arizona, California, Michigan, and Washington are reported to their respective NPCR state cancer registry for inclusion in those states' incidence data and are transmitted to CDC as part of the state's annual data submission.

In the 2023 data submission, CDC received information on more than 37.9 million invasive cancer cases, and 2.75 million in situ cases diagnosed during 1995 through 2021. In addition, 935,508 benign and borderline brain and central nervous system tumors were reported during 2004 through 2021. More than 1.6 million new invasive cancer cases are added each year.

In conjunction with the annual release of United States Cancer Statistics (USCS) data, NPCR evaluates each funded central cancer registry's data according to NPCR's standards for data completeness, timeliness, and quality. Registry data must meet these standards to be included in USCS data products.

The release of USCS data in products including the Data Visualizations tool and Public Use Databases exemplifies the progress achieved in creating a national system of cancer surveillance. Viewers are able to access locally relevant cancer data at the county and congressional district levels. Since these data are collected using national standards, users can compare information across geographic areas and over time. Data from state and county levels can be used to plan and evaluate cancer control programs, conduct research, and monitor cancer trends.

Partners such as the central cancer registry are crucial to the success of cancer surveillance in the United States. The efforts and achievements of many partner organizations contribute to USCS data products and the advances in cancer surveillance in the United States.

USCS data products include:

- A web-based data visualizations tool that displays USCS data, the official federal cancer statistics.
- Public use databases for researchers to analyze more than 35 million cancer cases.

- A public use data set of pre-calculated cancer incidence and death rates on CDC WONDER.
- Publications on cancer burden.
- A website designed to help guide and prioritize cancer control activities at the state and county level at State Cancer Profiles.
- A restricted-access dataset available to researchers through the National Center for Health Statistics Research Data Center.

Surveillance, Epidemiology, and End Results (SEER) Program

The National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) Program collects and publishes data on cancer incidence and survival from population-based cancer registries in 22 U.S. geographic areas. SEER registries provide complete coverage for metropolitan regions and special populations whose data are reported to their respective state registries funded by CDC's National Program of Cancer Registries.

The SEER Program continues to:

- Monitor the burden of cancer in the United States.
- Provide statistics on cancer incidence and survival in the SEER coverage area, and mortality in the United States provided by CDC's National Center for Health Statistics.
- Monitor cancer incidence trends in geographic and demographic population groups, including diverse racial and ethnic groups.
- Provide detailed information on trends in the extent of disease at diagnosis, therapy, and patient survival.
- Provide data for research.
- Promote studies measuring progress in cancer control and etiology.
- Provide specialty training in epidemiology, biostatistics, surveillance research, tumor registry methodology, operations, and management.
- Respond to public concerns and inquiries on cancer.
- Develop new statistical methods, models, and software for the analysis and presentation of national and small-area statistics.

The mortality data reported by SEER are provided by CDC's National Center for Health Statistics. The SEER Program issues a limited-use data set (formerly called the public use data file) for additional analyses by researchers and the public.

In addition to the data sets on the SEER website, NCI disseminates:

- A public use interactive website of pre-calculated cancer incidence rates called SEER*Explorer.
- A public use interactive tool to explore childhood, adolescent, and young adult cancer statistics called NCCR*Explorer.
- Cancer statistics fact sheets.

For information about NCI's additional tools and products, visit their Cancer Statistics page.

National Vital Statistics System (NVSS)

The nation's vital statistics are available from the National Vital Statistics System (NVSS), which is maintained by CDC's National Center for Health Statistics (NCHS). These vital statistics are provided through state-operated registration systems and are based on vital records filed in state vital statistics offices.

Recording vital events is the responsibility of the individual states and independent registration areas (District of Columbia, New York City, and five territories) in which the event occurs. Legal responsibility for the registration of vital events rests with the individual jurisdictions.

Through its Vital Statistics Cooperative Program, NCHS cooperates with state vital statistics offices to develop and recommend standard forms for data collection, model regulations, and procedures to ensure uniform reporting of the events monitored by the NVSS. Detailed annual data on births, deaths (including infant deaths), and fetal deaths are available for the United States and for states, counties, and other local areas.

The NCHS restricted use data set is obtained annually through NCHS's application process. Data variables include cause of death, age, race, Hispanic origin, sex, marital status, place of birth, residence of decedent, education level, and place of death.

Partners

Those crucial to the success of cancer registration and cancer surveillance in the United States include the American Cancer Society, the American College of Surgeons, the American Joint Committee on Cancer, the National Cancer Registrars Association, and the North American Association of Central Cancer Registries.

Incidence Data Sources

Incidence data are from the registries participating in the Centers for Disease Control and Prevention's (CDC's) National Program of Cancer Registries (NPCR) and the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) Program. NPCR funds and receives data from 50 cancer registries: 46 states, the District of Columbia, Puerto Rico, the Pacific Island Jurisdictions, and the U.S. Virgin Islands. SEER collects and publishes data from population-based cancer registries in 22 U.S. geographic areas. Data from central cancer registries that are supported by both NPCR and SEER are presented as reported to CDC in 2023.

How incidence data are collected

The primary source of data on cancer incidence is medical records. Staff at health care facilities abstract data from patients' medical records, enter it into the facility's own cancer registry (if it has one), and then send the data to the regional or state registry. Other data sources include physicians' offices, radiation facilities, freestanding surgical centers, and pathology laboratories. Both NPCR and SEER registries collect data using uniform data items and codes as documented by the North American Association of Central Cancer Registries (NAACCR). This standardization ensures that data items collected by the two federal programs are comparable.[1] [2] Information on primary site and histology was coded according to the International Classification of Diseases for Oncology, Third Edition (ICD-O-3)[3] and categorized according to the revised SEER recodes dated January 27, 2003, which define standard groupings of primary cancer sites. Beginning with 2010 diagnoses, cases are coded based on ICD-O-3 updated for hematopoietic codes based on WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (2008).

Reportable cases

Cancer is a mandatory, reportable disease for public health surveillance. NPCR and SEER cancer registries consider all incident cases with a behavior code of 2 (*in situ,* noninvasive) or 3 (invasive, primary site only) in the ICD-O-3 with the exception of in situ cancer of the cervix as reportable. Basal and squamous cell carcinomas of the skin are also excluded, with the exception of those on the skin of the genital organs.[3] Several cancers are coded as malignant in ICD-O-3 (beginning with 2001 diagnoses) that were not coded as malignant in ICD-O-2[3] and are noted as follows:

- Myelodysplastic syndrome (MDS) including refractory anemias (histology codes 9980, 9982–9984, 9989) are included in the "Miscellaneous" and "All Sites" categories.
- Chronic myeloproliferative disease (CMPD) including polycythemia vera and thrombocythemias (histology codes 9950, 9960–9962) are included in the "Miscellaneous" and "All Sites" categories.
- Papillary ependymomas (9393) and papillary meningiomas (9538)—cancers that occur in the central nervous system—are included in the "Brain and Central Nervous System" and "All Sites" categories.
- Some endometrial tumors (8931) are reported in the "Corpus and Uterus, NOS" and "All Sites" categories.

For comparisons with ICD-O-2 for cancers diagnosed prior to 2001, exclude all of the histology codes described above and listed as follows: 8931, 9393, 9538, 9950, 9960–9962, 9980, 9982–9984, 9989, 9990, 9991, 9992.[3]

Additional changes in ICD-O-3 apply to ovarian cancer: low malignant potential tumors (8442, 8451, 8462, 8472, 8473) of the ovary are no longer coded as malignant. Therefore, these cancers are not accounted for in the calculations of the incidence rate for ovarian cancer included in tables and figures. A footnote is provided as a reminder of this exclusion.

Pilocytic astrocytomas (9421) are also not coded as malignant in ICD-O-3, but these cancers are included in this report.

In situ bladder and breast cancers

In situ bladder cancers were recoded to invasive bladder cancers because the information needed to distinguish between in situ and invasive bladder cancers is not always available or reliable. Counts and rates for in situ breast cancer cases among women are presented; these are reported separately and are not included in counts or rates for the "All Types of Cancer" category.

Unknown sex, age, or race

Non-reportable cancers and cancers in patients of unknown sex or age were omitted from all calculations, but cancers in patients of unknown race were included in the "All Races" category.

Childhood cancer

Incidence data on childhood cancer are published in two formats:

- The first is according to the SEER modification of the third edition of the International Classification of Childhood Cancer. The ICCC-3 is based on ICD-O-3/IARC 2017 classification of Tumors of Haematopoietic and Lymphoid tissues.[4] The ICCC presents childhood cancers in 12 groups classified primarily by morphology. The SEER modification, which affects the classification of nervous system and bone tumors, was chosen for compatibility with other published data on rates of childhood cancer in the United States.
- The second format is according to the SEER site recode, which is based primarily on cancer site; the incidence data are presented in this format to make them comparable with published mortality data. This format allows the incidence data for childhood cancers to be categorized in the same groups as adult cancers. Although these groupings are not as appropriate for children as they are for adults, they are necessary to allow comparisons between childhood incidence and childhood mortality.

Nonmalignant brain and CNS tumors

Incidence data on nonmalignant primary brain and central nervous system (CNS) tumors are available in the U.S. Cancer Statistics Data Visualizations tool. Cancer registries began collecting information on nonmalignant brain and CNS tumors beginning with 2004 diagnoses. Data collection of these tumors is in accordance with Public Law 107-260, the Benign Brain Tumor Cancer Registries Amendment Act, which mandates that NPCR registries collect data on all brain and CNS tumors with a behavior code of 0 (benign) and those with a behavior code of 1 (borderline), in addition to in situ and malignant.[5] SEER registries voluntarily agreed to incorporate registration of these tumors in their standard practices.[5]

Effects of events on cancer incidence data

Effect of Hurricanes Katrina and Rita on presenting cancer incidence data

The population of many counties along the Gulf Coast of Louisiana, Alabama, Mississippi, and Texas were displaced in the fall of 2005 by Hurricanes Katrina and Rita, resulting in incomplete case ascertainment for the latter half of the year. For these states, state- and county-level incidence rates were calculated based upon the data submitted to CDC.

Effect of Hurricane Maria on presenting Puerto Rico's cancer incidence data

Puerto Rico's 2017 incidence counts and corresponding rates are based on the first 6 months of the reported data coupled with half of the population estimate (January to June 2017). Cases with an unknown month of diagnosis were also included. To account for the population shift that occurred due to Hurricane Maria in September 2017, data from July to December 2017 are excluded. This population shift may have resulted in incomplete case ascertainment for the latter half of the year.

The population denominators were adjusted by dividing the US Bureau of the Census's July 1, 2017 (vintage 2021) Puerto Rico population estimate in half.

Effect of COVID-19 on cancer incidence data for diagnosis year 2020

In March 2020, the World Health Organization declared COVID-19 a pandemic. Soon after, stay-at-home orders, business and school shutdowns, and travel advisories were implemented in the United States to prevent the spread of COVID-19. Additionally, some health care systems reduced access to routine care. These measures, along with concerns about getting COVID-19, interrupted cancer screening, diagnosis, and care, as people postponed or deferred health care visits, particularly from March to May 2020.

The 2023 data submission includes new cancer cases diagnosed in 2020 and 2021, the first and second years of the COVID-19 pandemic. The missed cancer diagnoses resulting from disruptions in health services caused by the pandemic may have contributed to an observed decline in incidence for most cancer sites in 2020.[6,7] The numbers of new cases diagnosed in 2021 are still a little lower than expected for some cancer types but have returned to pre-pandemic counts for other cancer types. Caution must be taken when examining trends to avoid incorrect interpretations of the effect of cancer prevention and early detection efforts. Observed downward trends may be due largely to the lower observed incidence in 2020.[6] CDC and NCI include the 2020 incidence rates in statistical reports and graphics, but do not include them in joinpoint models. The 2021 incidence data will be included in statistical reports and joinpoint models (not included in Data Visualizations tool). JoinPoint software allows researchers to exclude incidence data for 2020, 2021, or both years from trend analyses. Exclude 2020 data for incidence trend analyses, but 2021 data can be included in reports and trend analyses. Learn more about the impact of COVID-19 on cancer incidence trends.

References

1. National Cancer Institute. SEER Program Coding and Staging Manual 2023. Accessed March 19, 2024.
2. Thornton ML, (ed). *Standards for Cancer Registries Vol II: Data Standards and Data Dictionary, Version 24, 25th edition*. Springfield, Ill: North American Association of Central Cancer Registries, July 2023, revised July 2023, August 2023, October 2023.
3. Fritz A, Percy C, Jack A, et al. *International Classification of Diseases for Oncology, Third Edition*. Geneva, Switzerland: World Health Organization; 2000.

4. World Health Organization. Hematopoietic codes based on *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues* (2008).

5. McCarthy BJ, Kruchko C, and the Central Brain Tumor Registry of the United States. Consensus conference on cancer registration of brain and central nervous system tumors. *Neuro Oncol.* 2005;7(2):196–201.

6. Centers for Disease Control and Prevention. Highlights from 2021 Cancer Incidence with Comparisons to Previous Years. Centers for Disease Control and Prevention, US Department of Health and Human Services; 2024.

7. Mariotto AB, Feuer EJ, Howlader N, Chen H, Negoita S, Cronin KA. Interpreting cancer incidence trends: challenges due to the COVID-19 pandemic. *J Natl Cancer Inst.* 2023;115(9):1109–1111.

Mortality Data Sources

How mortality data are collected

Cancer mortality statistics are based on information from all death certificates filed in the 50 states, the District of Columbia, and Puerto Rico and processed by CDC's National Center for Health Statistics (NCHS) National Vital Statistics System (NVSS). The cancer mortality data were compiled in accordance with World Health Organization (WHO) regulations, which specify that member nations classify and code causes of death in accordance with the current revision of the International Classification of Diseases (ICD). Effective with deaths that occurred in 1999, the United States began using the tenth revision of this classification (ICD–10).[1 2]

Rules for coding the cause(s) of death may require modification when evidence suggests that such modifications will improve the quality of cause-of-death data. Before 1999, such modifications were made only when a new revision of the ICD was implemented. A process for updating the ICD that allows for mid-revision changes was introduced with ICD-10. Minor changes may be implemented every year, while major changes may be implemented every three years. These updates do not have a significant effect on the data in the U.S. Cancer Statistics Data Visualizations tool.

The ICD not only details disease classification but also provides definitions, tabulation lists, the format of the death certificate, and the rules for coding cause of death. Cause-of-death data presented in the U.S. Cancer Statistics Data Visualizations tool were coded by procedures outlined in annual issues of the NCHS Instruction Manuals.

Underlying cause of death

In the U.S. Cancer Statistics Data Visualizations tool, tabulations of cause-of-death statistics are based solely on the underlying cause of death, which is defined by WHO as "the disease or injury that initiated the train of events leading directly to death, or the circumstances of the accident or violence that produced the fatal injury."[1] The underlying cause of death is selected from the conditions entered by the physician in the cause-of-death section of the death certificate. Generally, more medical information is reported on death certificates than is reflected directly in the underlying cause of death.[3 4]

Kaposi sarcoma

Because the vast majority of Kaposi sarcoma cases develop in association with human immunodeficiency virus (HIV) infection and acquired immunodeficiency syndrome (AIDS), HIV/AIDS is listed as the underlying cause of death. Therefore, Kaposi sarcoma death rates are not included.

Cancer site groups

For consistency with the data on cancer incidence, the cancer sites in mortality data were grouped according to the revised SEER recodes dated March 1, 2018. Because NCHS uses different groupings for some sites, the death rates in this report may differ slightly from those published by NCHS. In addition, under the ICD, there are differences in mortality and incidence coding. For example, in ICD-10, mesothelioma deaths are coded by anatomic site whereas in ICD-O-3, mesothelioma incidence is coded by morphology, regardless of anatomic site.

Mortality data submission process

Unlike incidence data, mortality data for a calendar year are not updated after the final data file is released. All states, the District of Columbia, and Puerto Rico submitted their 2023 mortality data to NCHS electronically. Mortality data for the entire United States refer to deaths that occurred within the United States. Data for geographic areas are grouped by the decedent's place of residence.

References

1. Xu J, Murphy SL, Kochanek KD, Arias E. Deaths: Final data for 2019. *Natl Vital Stat Rep.* 2021;69(13).
2. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision.* Geneva, Switzerland: World Health Organization; 1992.
3. Chamblee RF, Evans MC. TRANSAX: the NCHS system for producing multiple cause-of-death statistics, 1968–78. *Vital Health Stat 1.* 1986;(20):1–83.
4. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. *A J Epidemiol.* 1986;124(2):161–179.

Population Denominator Data Sources

The population estimates for the denominators of incidence and death rates are race-specific, ethnicity-specific, age-specific, and sex-specific county population estimates aggregated to the state or metropolitan-area level.

U.S. Census Bureau population estimates

The U.S. Bureau of the Census (Census Bureau) produces estimates of the U.S. population annually (each new series is called a "Vintage"). In general, July 1 population estimates are used because these estimates are thought to reflect the average population of a geographic area for a calendar year. The Census Bureau has released intercensal population estimates through year 2009 and postcensal estimates for years 2010 and later (Census intercensal estimates for years 2010–2019 are scheduled to be released in fall 2024).

The Census Bureau uses six "single-race" race categories (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, White, More than one race) as specified in the 1997 Office of Management and Budget (OMB) standards for the collection of data on race and ethnicity. CDC's National Center for Health Statistics released "bridged-race" population estimates through 2020 for use in calculating vital rates. These estimates result from bridging the race categories specified in the 1997 OMB standards to the race categories (Asian or Pacific Islander, Black or African American, American Indian or Alaska Native, White) specified in the 1977 standards.

NCI modifications to population estimates

The county population estimates that are incorporated into the National Cancer Institute's (NCI's) SEER*Stat software are a slight modification of the annual time series of July 1 county population estimates (by age, sex, race, and Hispanic origin) produced by the Census Bureau, in collaboration with CDC's National Center for Health Statistics, with support from NCI. The NCI population datasets use bridged-race categories. Under agreements with NCI, Woods & Poole developed bridged race intercensal population estimates for 2010–2019 and the Census Bureau developed bridged race postcensal population estimates for 2020–2022.

NCI's modifications to the population estimates are documented. Several modifications pertain to the grouping of counties needed to assure the compatibility of all incidence, mortality, and population data sets.

Population estimates for Hawaii

Another modification only affects population estimates for the state of Hawaii. Based on concerns that the native Hawaiian population has been vastly undercounted in previous censuses, the Epidemiology Program of the Hawaii Cancer Research Center recommended an adjustment to the populations for its state. The "Hawaii adjustment" to the Census Bureau's estimates has the net result of reducing the estimated White population and increasing the estimated Asian and Pacific Islander population for the state. The estimates for the total population, Black population, and American Indian and Alaska Native population in Hawaii are not modified.

Population estimates for Puerto Rico

Population estimates used in the calculation of rates for Puerto Rico are:

- Sex-specific and age-specific.
- Obtained from the U.S. Census Bureau.
- Not available by race or ethnicity.

Modifications for special circumstances

Effects of Hurricanes Katrina and Rita

The populations of many counties along the Gulf Coast of Louisiana, Alabama, Mississippi, and Texas were displaced in the fall of 2005 by Hurricanes Katrina and Rita. The population estimates were adjusted to account for the displacement of people in these states. The national total population estimates are not affected by these adjustments.

The majority of the evacuees from Hurricanes Katrina and Rita relocated to Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, Tennessee, or Texas. The evacuee population was included in the 2005 incidence rates, since all of the relocation states met the U.S. Cancer Statistics publication criteria.

Effects of Hurricane Maria

Similarly, to minimize the effect of Hurricane Maria in Puerto Rico in September 2017, modified Puerto Rico population estimates obtained from the U.S. Census Bureau are used to calculate cancer incidence rates for Puerto Rico for 2017. The population denominators were adjusted by dividing the US Census Bureau's July 1, 2017 (vintage 2021) Puerto Rico population estimate in half.

Population estimates used in rate calculations

Cancer incidence rates for 1999–2021 were calculated using the NCI-modified bridged-race population estimates for denominators.

Cancer death rates for 2018–2022 were calculated using U.S. Census Bureau single-race population estimates for denominators. Cancer death rates for 1999–2017 overall trend analyses were calculated using NCI-modified bridged-race population estimates for denominators.

Screening, HPV Vaccination, and Risk Factor Data Sources

Data on screening and risk factors

Data on cancer screening and risk factors are based on information obtained from the Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS is a system of ongoing state-based health-related telephone surveys that collect data on health-related risk behaviors, chronic health conditions, and use of preventive services.

The BRFSS surveys adults in all 50 states, the District of Columbia, and three U.S. territories. The BRFSS completes more than 400,000 interviews each year.

The BRFSS field operations are managed by state health departments that follow protocols adopted by the states, with technical assistance from CDC. The data are transmitted to CDC for editing, processing, weighting, and analysis. CDC provides an edited and weighted data file to each participating health department for each year of data collection. CDC also prepares summary reports of geographic area-specific data.

Data on HPV vaccinations

Data on human papillomavirus (HPV) vaccinations are collected through the National Immunization Survey-Teen (NIS-Teen). The NIS-Teen is a random-digit-dialed survey of parents or guardians of teens who are 13 to 17 years old. The telephone survey is followed by a questionnaire mailed to vaccination providers to obtain the teen's vaccination history.

The national sample contains more than 20,000 teens with adequate vaccination coverage data reported by their health care providers (adequate provider data). Vaccination coverage estimates are based on provider-reported vaccination histories.

Registries That Met U.S. Cancer Statistics Publication Criteria

Publication criteria were assessed based on data submitted to CDC's National Program of Cancer Registries and the National Cancer Institute's Surveillance, Epidemiology, and End Results Program in 2023. Criteria must be met for all diagnosis years for the combined 2017 to 2021 data. Segments of a registry's data may be suppressed when that information is found to be unstable or does not meet the publication criteria.

Population coverage by diagnosis years

The following registries met U.S. Cancer Statistics publication criteria during each year it has been published:

- **2017 to 2021:** All registries except Indiana met the publication criteria for all years during this 5-year period. Counts and rates cover 98% of the U.S. population.
- **1999 to 2021:** All registries met the publication criteria for all years during 1999–2021 except Mississippi (2002) and Indiana (2020–2021). Data are not available for Mississippi (1999–2001) and South Dakota (1999–2000). Counts and rates cover approximately 99.6% of the U.S. population.
- **2021:** All registries except Indiana met the publication criteria. Counts and rates cover 98% of the U.S. population.
- **2020:** All registries except Indiana met the publication criteria. Counts and rates cover 98% of the U.S. population.
- **2019:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2018:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2017:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2016:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2015:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2014:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2013:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2012:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2011:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2010:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2009:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2008:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2007:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2006:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2005:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2004:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2003:** All registries met the publication criteria. Counts and rates cover 100% of the U.S. population.
- **2002:** All registries except Mississippi met the publication criteria. Counts and rates cover approximately 99% of the U.S. population.

- **2001:** All registries met the publication criteria; data are not available for Mississippi. Counts and rates cover approximately 99% of the U.S. population.
- **2000:** All registries met the publication criteria; data are not available for Mississippi and South Dakota. Counts and rates cover approximately 99% of the U.S. population.
- **1999:** All registries met the publication criteria; data are not available for Mississippi and South Dakota. Counts and rates cover approximately 99% of the U.S. population.

Incidence and Death Rates

Crude rates are helpful in determining the cancer burden and specific needs for services for a given population, compared with another population, regardless of size. Crude rates are calculated as follows:

- Crude and age-specific incidence rates equal the total number of new cancer cases diagnosed in a specific year in the population category of interest, divided by the at-risk population for that category and multiplied by 100,000 (cancers by primary site) or by 1 million (International Classification of Childhood Cancer [ICCC] groupings of childhood cancers).
- Crude and age-specific death rates equal the total number of cancer deaths during a specific year in the population category of interest, divided by the at-risk population for that category and multiplied by 100,000.

Crude rates and age-adjusted rates

Crude rates are influenced by the underlying age distribution of the state's population. For example, if two states have the same age-adjusted rates, the state with the relatively older population generally will have higher crude rates because incidence or death rates for most cancers increase with increasing age.

The age distribution of a population (the number of people in particular age groups) can change over time and can be different in different geographic areas. Age-adjusting the rates ensures that differences in incidence or deaths from one year to another, or between one geographic area and another, are not due to differences in the age distribution of the populations being compared.

2000 U.S. standard population age groups

The 2000 U.S. standard population[1 2] is used to age-adjust the rates in this report. In the U.S. Cancer Statistics Data Visualizations tool, the 2000 U.S. standard population is based on the proportion of the 2000 population in 19 age groups (younger than 1 year, 1 to 4 years, 5 to 9 years, 10 to 14 years, 15 to 19 years, and so on through 85 years and older) except for Puerto Rico. In Puerto Rico, the two youngest age groups are combined, so the standard population is based on 18 age groups (0 to 4 years, 5 to 9 years, 10 to 14 years, and so on). The proportions of the 2000 population in these age groups serve as weights for calculating age-adjusted incidence and death rates. The National Center for Health Statistics (NCHS) regularly evaluates the population standard and currently recommends using the 2000 U.S. standard population for calculating age-adjusted rates.

Cancer death rates in the Data Visualizations tool may differ slightly from those published by the National Center for Health Statistics (NCHS) because NCHS uses the age groups as recommended by the U.S. Department of Health and Human Services to adjust death rates.

The 2000 U.S. standard population weights are not race- or sex-specific, so they do not adjust for differences in race or sex distribution between geographic areas or populations being compared. They do, however, provide the basis for adjusting for differences in the age distributions across groups defined by sex, race, ethnicity, geography, or other categories.

The 2000 U.S. standard population weights used for this report are based on single years of age from the Census P25-1130 series estimates of the 2000 U.S. population. Populations for single years of age are summed

to form the age groups. These standard weights are used to compute age-adjusted incidence and death rates by the method of direct standardization as implemented in the National Cancer Institute's SEER*Stat software.

Ideally, crude, age-adjusted, and age-specific rates are used to plan for population-based cancer prevention and control interventions.

References

1. Anderson RN, Rosenberg HM. Report of the Second Workshop on Age Adjustment. *Vital Health Stat 4*. 1998;(30):I–VI, 1–37.

2. Anderson RN, Rosenberg HM. Age standardization of death rates: implementation of the year 2000 standard. *Natl Vital Stat Rep*. 1998;47(3):1–16, 20.

Confidence Intervals

Width of confidence intervals

The width of a confidence interval depends on the amount of variability in the data. Narrow confidence intervals tend to imply greater certainty in the estimate, while wide confidence intervals tend to imply more variability in the data and could mean there is less certainty.

Sources of variability include the underlying occurrence of cancer as well as uncertainty about when the cancer is diagnosed, when a death from cancer occurs, and when the data about the cancer are sent to the registry or state health department.

In any year when large numbers of a particular cancer are diagnosed or large numbers of cancer patients die, the effects of random variability are small and the confidence interval would likely be narrow. With rare cancers, however, the rates are small and the chance occurrence of more or fewer cases or deaths in a year can affect those rates markedly. Under these circumstances, the confidence interval will be wide to indicate uncertainty or instability in the cancer rate.

The Poisson process

To estimate the extent of this uncertainty, a statistical framework is applied.[1] The standard model used for rates for vital statistics is the Poisson process,[2] which assigns more uncertainty to rare events relative to the size of the rate than it does to common events.

Parameters are estimated for the underlying disease process. For this report, we estimated a single parameter to represent the incidence rate and its variability. Of note, the Poisson model can estimate separate parameters that represent contributions to the rate from various risk factors, the effects of cancer control interventions, and other attributes of the population risk profile in any year.

Modified gamma intervals

The Data Visualizations tool uses confidence intervals that are expected to include the true underlying rate 95% of the time. The confidence intervals are modified gamma intervals[3] computed using SEER*Stat. The modified gamma intervals are more efficient than the gamma intervals of Fay and Feuer[4] in that they are less conservative while still retaining the nominal coverage level.

Various factors such as population heterogeneity can sometimes lead to "extra-Poisson" variation in which the rates are more variable than would be predicted by a Poisson model. No attempt was made to correct for this. In addition, the confidence intervals do not account for systematic (in other words, nonrandom) biases in the incidence rates.

Considerations when comparing rates

Using overlapping confidence intervals to determine significant differences between two rates presented in the Data Visualizations tool is discouraged. The practice fails to detect significant differences more frequently than standard hypothesis testing.[5]

Another consideration when comparing differences between rates is their public health importance. For some rates presented in the Data Visualizations tool, numerators and denominators are large and standard errors are therefore small. This results in statistically significant differences that may be too small to be important for decisions related to population-based public health programs.

References

1. Särndal C-E, Swennson B, Wretman J. *Model-Assisted Survey Sampling.* New York (NY): Springer-Verlag; 1992.

2. Brillinger DR. The natural variability of vital rates and associated statistics. *Biometrics.* 1986;42(4):693–734.

3. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res.* 2006;15(6):547–569.

4. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med.* 1997;16(7):791–801.

5. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *Am Stat.* 2001;55(3):182–186.

Stage at Diagnosis

Summary Stage

Stage measures how far a cancer has spread from its origin. CDC's National Program of Cancer Registries and the National Cancer Institute's Surveillance, Epidemiology, and End Results Program use the Summary Stage staging system. Summary Stage characterizes invasive cancers as localized, regional, or distant.[1]

- Localized cancer is confined to the primary site.
- Regional cancer has spread directly beyond the primary site (regional extension) or to regional lymph nodes.
- Distant cancer has spread to other organs (distant extension) or remote lymph nodes.[1]
- Some cancers are unstaged, or the stage is unknown or unspecified.

Stage categories are different for two cancer sites:

- For brain and central nervous system tumors, the regional and distant categories have been combined.
- For urinary bladder tumors, *in situ* primaries are included as a category.

In the U.S. Cancer Statistics Data Visualizations tool, stage is classified using the *Merged Summary Stage* variable. *Merged Summary Stage* data are suppressed for testis cases due to a data quality issue.

How stage distribution data are presented

Stage distribution data are presented as case counts and percentages for two groups:

- The first group includes localized, regional, distant, and unstaged cases. Including unstaged cases helps quantify the amount of missing data and enables comparisons with other studies using this categorization. However, including unstaged cases will underestimate the percentages of the other stage categories.

- The second group includes only the known stage categories (localized, regional, and distant). Excluding unknown stage provides better estimates of the stage category percentages.

Frequencies and percentages are suppressed for groups with fewer than 16 cases. In addition, complementary cell suppression is done to suppress data for both sexes combined if data are suppressed for one sex. For more information, see Suppression of Rates and Counts.

Reference

1. Ruhl JL, Callaghan C, Schussler N (eds.) *Summary Stage 2018: Codes and Coding Instructions.* National Cancer Institute, Bethesda, MD, 2023.

Relative Cancer Survival

Cancer survival estimates can be used to support the needs of people who have cancer now or had cancer in the past, estimated to be 14 million in 2021.[1]

Definition and calculation of relative cancer survival

Relative cancer survival measures the proportion of people with cancer who will be alive at a certain time after diagnosis, given that they did not die from something other than their cancer during that period of time. Relative cancer survival is defined as the ratio of the observed all-cause survival in a group of individuals with cancer to the expected all-cause survival of a similar group of individuals who do not have cancer.[2]

Because the expected survival of individuals who do not have cancer is difficult to obtain, it is often approximated by the expected all-cause survival of the general population. This is a reasonable approximation because cancer deaths are generally a negligible proportion of all deaths. Thus, the relative cancer survival is calculated as the observed all-cause survival in a group of individuals with cancer divided by the expected all-cause survival of the general population. To learn more on this topic, visit Measures of Cancer Survival.

How relative cancer survival rates are calculated

Cancer incidence data submitted to CDC's National Program of Cancer Registries (NPCR) in the 2023 data submission period were used to create a data set in SEER*Stat for this analysis.[3] The data set included data from 43 NPCR-funded central cancer registries that:

- Met the United States Cancer Statistics (USCS) publication criteria for all years 2014 through 2020.
- Conducted linkage with the National Death Index, active patient follow-up for all years 2014 through 2020, or both.

These registries include Alabama, Alaska, Arizona, Arkansas, California, Colorado, Delaware, Florida, Georgia, Idaho, Illinois, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, Wisconsin, and Wyoming. These data cover 92% of the U.S. population.

Cases from these registries were included in the analysis if:

- The case was an invasive cancer diagnosed from 2014 through 2020. Cases diagnosed in 2021 do not have adequate follow-up time to be included in the analysis.
- The patient's age was known and was 0 through 99 years.
- The patient's sex was known.
- The case was not identified solely on the basis of a death certificate or autopsy.

Analytic methods

Survival time

Survival time in months for each case was calculated:

- Date of the start of follow-up (month, day, and year) was set to the date of diagnosis.
- Date of the last follow-up (month, day, and year) was set to:
    - The date of death if the case was matched to the state death files or the National Death Index.
    - The date of last contact if the case was actively followed.
    - December 31, 2020 if the case was not matched to the state death files or the National Death Index and was not actively followed.
- If the day or month was missing for the date of diagnosis, date of death, or date of last contact, the full date was imputed using a standard algorithm.[4]
- Patients who lived beyond age 99 were excluded.

Observed all-cause survival for people with cancer

Observed all-cause survival by sex, race, and ethnicity (all races, non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, non-Hispanic Asian and Pacific Islander, and Hispanic) for individuals with any cancer and for individuals with 25 common cancer sites was then calculated using the actuarial life table method.[5]

Multiple primaries

Cases with multiple primary cancers were included in the dataset, although only the first primary cancer during the inclusion period was included in calculating relative survival for all cancer sites combined. If a patient had multiple primary cancers of different sites, each cancer was included in calculating cancer-specific relative survival. If a patient was diagnosed with multiple primary cancers of the same site at the same age, only the first primary cancer was included in calculating relative survival for that cancer site, and only one record per person will contribute to any life page (strata in a data visualization query).[6]

Expected all-cause survival for the general population

Expected all-cause survival for the general population was obtained using annual US life tables provided by the National Center for Health Statistics and modified by SEER. The general population was grouped by sex, race and ethnicity (all races, non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, non-Hispanic Asian and Pacific Islander, and Hispanic), geography (state and county), and socioeconomic status. The life tables were embedded in SEER*Stat. See Expected Survival Life Tables for more information.

Relative cancer survival

The Ederer II method[2] was used to calculate relative survival for all cancer sites combined and for 25 common cancer sites by sex, race and ethnicity (all races, non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, non-Hispanic Asian and Pacific Islander, and Hispanic), and age group (younger than 45 years, 45 to 54 years, 55 to 64 years, 65 to 74 years, and 75 years or older). See Measures of Cancer Survival for more information.

Relative cancer survival is presented by:

- State for all cancer sites combined and for 25 common sites by sex and by race and ethnicity.
- Stage for 24 common sites (testis excluded) by sex and by race and ethnicity.
- Age at the national level.

The quality and completeness of individual data items used in this analysis are discussed in a study by Wilson and others.[7]

References

1. U.S. Cancer Statistics Data Visualizations Tool. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, released in June 2024.

2. Cho H, Howlader N, Mariotto AB, Cronin KA. Estimating relative survival for cancer patients from the SEER Program using expected rates based on Ederer I versus Ederer II method. Surveillance Research Program, National Cancer Institute; 2011. Technical Report #2011-01.

3. National Program of Cancer Registries SEER*Stat Database: NPCR Survival Analytic file 2001–2020 (43 NPCR central cancer registries). United States Department of Health and Human Services, Centers for Disease Control and Prevention. Released June 2024, based on the 2023 submission.

4. Johnson CJ, Weir HK, Yin D, Niu X. The impact of patient follow-up on population-based survival rates. *J Registry Manag.* 2010;37(3):86–103.

5. Lee ET. Life-table analysis. In: *Statistical Methods for Survival Data Analysis,* 2nd ed. New York, NY: John Wiley & Sons, 1992: 78–100.

6. Brenner H, Hakulinen T. Patients with previous cancer should not be excluded in international comparative cancer survival studies. *Int J Cancer.* 2007;121(10):2274–2278.

7. Wilson RJ, O'Neil ME, Ntekop E, Zhang K, Ren Y. Coding completeness and quality of relative survival-related variables in the National Program of Cancer Registries Cancer Surveillance System, 1995–2008. *J Registry Manag.* 2014;41(2):65–71.

Cancer Prevalence

Definition and calculation of cancer prevalence

Prevalence is the number of people with a specific disease or condition in a given population at a specific time. This measure includes both newly diagnosed and pre-existing cases of the disease. It is different from incidence, because incidence measures only the number of newly diagnosed cases in a given population at a specific time.

There are different types of prevalence. For example:

- Annual prevalence is the number of people with the disease at any time during a year.
- Period prevalence is the number of people with the disease at any time during a specified number of years, such as the last 10 years.
- Limited-duration prevalence is the number of people alive on a certain day who were diagnosed with the disease during a specified number of years (such as the last 5 or 20 years).

How cancer prevalence is calculated

Cancer incidence data submitted to CDC's National Program of Cancer Registries (NPCR) in the 2023 data submission period were used to create a data set in SEER*Stat for this analysis.[1,2] The data set included data from 43 NPCR central cancer registries that:

- Met the United States Cancer Statistics (USCS) publication criteria for all years 2001 through 2020.
- Conducted linkage with the National Death Index, active patient follow-up for all years 2001 through 2020, or both.

These registries include Alabama, Alaska, Arizona, Arkansas, California, Colorado, Delaware, Florida, Georgia, Idaho, Illinois, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, Wisconsin, and Wyoming. These data cover 92% of the U.S. population.

Cases from these registries were included in the analysis if:

- The case was an invasive cancer diagnosed from 2001 through 2020.
- The patient's age was known and was 0 through 99 years.
- The patient's sex was known.
- The case was not identified solely on the basis of a death certificate or autopsy.

Because NPCR data are available from 2001, 20-year limited-duration prevalence estimates are included in addition to 5-year estimates.

Calculation of limited-duration prevalence

Limited-duration prevalence is the number of people alive on a certain day who were diagnosed with the disease during a specified number of years (such as the last 5 or 20 years).

In this report, the limited-duration prevalence was calculated using SEER*Stat software.[2] It estimates, among the people diagnosed with cancer in the last 5 or 20 years, the proportion who were still alive on January 1, 2021.[1]

- The date of the start of follow-up (month, day, and year) was set to the date of diagnosis.
- The date of the last follow-up (month, day, and year) was set to:
    - The date of the last contact if the case was actively followed.
    - The date of death if the case was matched to the state death files or the National Death Index.
    - Cases not matched to the state death files or the National Death Index were presumed to be alive on the prevalence date.

Multiple primaries

For patients diagnosed with multiple tumors, prevalence calculations include the first tumor of each cancer type in the previous 5 or 20 years.

For example, a woman was diagnosed with thyroid cancer 9 years ago and breast cancer 3 years ago:

- The thyroid cancer would contribute to the 20-year limited-duration prevalence estimates for all cancer sites and for thyroid cancer.
- The breast cancer would contribute to the 5-year limited-duration prevalence estimate for all cancer sites and both the 5-year and 20-year estimates for breast cancer.
- The breast cancer would not contribute to the 20-year limited-duration prevalence estimate for all cancer sites because the woman is already counted in this estimate for thyroid cancer.

NPCR prevalence proportions

NPCR prevalence proportions were calculated for each combination of age, sex, and race and ethnicity group. For this section of the report, race and ethnicity were categorized as:

- Non-Hispanic White. Cases with unknown race were combined with White race.
- Non-Hispanic Black.
- Non-Hispanic, Indian Health Service-linked American Indian and Alaska Native.
- Non-Hispanic Asian and Pacific Islander.
- Hispanic.

Cancer prevalence counts for the U.S. population

Cancer prevalence counts for the U.S. population as of January 1, 2021, were estimated by multiplying the age-, sex-, and race and ethnicity-specific NPCR prevalence proportions by the corresponding U.S. population estimates. The U.S. population estimates are based on the average of the 2020 and 2021 population estimates from the U.S. Census Bureau. The sum of the counts by race and ethnicity was used to estimate the U.S. cancer prevalence counts for all races combined.[3] Cancer prevalence counts and percentages for each of the 43 states by sex and by race and ethnicity were estimated directly in SEER*Stat.

Prevalence percentage

Prevalence percentage is the percentage of the population alive with cancer. The U.S. prevalence percentage estimates are based on the states included in the analysis.

References

1. National Program of Cancer Registries SEER*Stat Database: NPCR Prevalence Analytic file 2001–2020 (43 NPCR central cancer registries). United States Department of Health and Human Services, Centers for Disease Control and Prevention. Released June 2024, based on the 2023 submission.

2. Surveillance Research Program, National Cancer Institute SEER*Stat software version 8.4.

3. Gail MH, Kessler L, Midthune D, Scoppa S. Two approaches for estimating disease prevalence from population-based registries of incidence and total mortality. *Biometrics.* 1999;55(4):1137–1144.

Screening, HPV Vaccination, and Risk Factor Prevalence Estimates

Screening and risk factors

Healthy behaviors such as being physically active, avoiding tobacco, limiting the amount of alcohol you drink, and getting cancer screening tests when recommended may prevent or help manage cancer.[1]

Monitoring health risk behaviors and the use of health care is fundamental to the development of effective public health programs and policies at the state and local levels.[2]

Cancer registries do not routinely collect information on health risk behaviors. So these data are obtained from the Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS is the nation's premier system of state-based health-related telephone surveys. The surveys collect data about adults' health-related risk behaviors, chronic health conditions, and use of preventive services. The BRFSS collects data in all 50 states, the District of Columbia, and three U.S. territories., and completes more than 400,000 interviews each year.

Prevalence is the measured or estimated percentage of people with an attribute or disease during a specific time period. The prevalence estimates are from the BRFSS core survey at the state level, as well as model-based prevalence estimates for all counties in the United States.

Age-adjusted prevalence is the measured or estimated percentage of people with an attribute or disease during a specific time period, standardized by direct method to the age distribution of the:

- U.S. 2000 standard million population for state-level estimates.
- Year 2000 U.S. population, distribution 9 for county estimates.[3]

Crude and age-adjusted prevalence are displayed for:

- Female breast, cervical, and colorectal cancer screening.
- Unhealthy behaviors including physical inactivity, poor nutrition, alcohol consumption, tobacco use, and obesity.

Estimates in the U.S. Cancer Statistics Data Visualizations tool may differ slightly from those published by the National Center for Health Statistics (NCHS) because NCHS uses the age groups recommended by the U.S. Department of Health and Human Services in its adjustment of death rates.

The 2000 U.S. standard population weights are not race- or sex-specific, so they do not adjust for differences in race or sex distribution between geographic areas or populations being compared. They do, however, provide the basis for adjusting for differences in the age distributions across groups defined by sex, race and ethnicity, geography, or other categories.

Data are suppressed if there are fewer than 50 respondents in a specific category such as sex, race, or ethnicity.

Small area estimates (county level)

County-level estimates are calculated using an innovative peer-reviewed multilevel regression and poststratification approach.[4] The primary data sources are BRFSS and the Census 2010 population. You can learn details about the methodology.

The following list shows the measures displayed in the U.S. Cancer Statistics Data Visualizations tool. For measures included in the PLACES website, links to their definitions are also provided.

Cancer screening

- Up-to-date with colorectal cancer screening among adults aged 50 to 75 years
- Mammography use among women aged 50 to 74 years
- Cervical cancer screening among women aged 21 to 65 years

Physical activity

- No leisure-time physical activity among adults aged ≥18 years
- 150 minutes or more of aerobic physical activity per week among adults aged ≥18 years

Nutrition

- Consumed fruits less than once a day among adults aged ≥18 years
- Consumed vegetables less than once a day among adults aged ≥18 years

Alcohol use

- Binge drinking among adults aged ≥18 years

Tobacco use

- Current smoking among adults aged ≥18 years
- Ever smoking among adults aged ≥18 years
- Never smoking among adults aged ≥18 years

Obesity

- Obesity among adults aged ≥18 years

Measures not available in PLACES

The following measures are not available in PLACES and are defined below. For more information, see the Calculated Variables in the 2021 Data File of the Behavioral Risk Factor Surveillance System.

150 minutes or more of aerobic physical activity per week

Demographic group: Adults aged 18 years or older.

Numerator: Adults aged 18 years or older who report 150 or more minutes of aerobic physical activity per week.

Denominator: Number of adults aged 18 years or older who reported information about any or no physical activity in the past month (excluding those who refused to answer, had a missing answer, or answered "don't know/not sure").

Measures of frequency: Annual prevalence: crude and age-adjusted (standardized by the direct method to the year 2000 standard U.S. population, distribution 9)[3] with 95% confidence intervals and by demographic characteristics when feasible.

Time period of case definition: Past month.

Consumed fruits less than once a day

Demographic group: Adults aged 18 years or older.

Numerator: Adults aged 18 years or older who report consuming fruits less than once per day.

Denominator: Number of adults aged 18 years or older who reported information about fruit consumption (excluding those who refused to answer, had a missing answer, or answered "don't know/not sure").

Measures of frequency: Annual prevalence: crude and age-adjusted (standardized by the direct method to the year 2000 standard U.S. population, distribution 9)[3] with 95% confidence intervals and by demographic characteristics when feasible.

Time period of case definition: Per day.

Consumed vegetables less than once a day

Demographic group: Adults aged 18 years or older.

Numerator: Adults aged 18 years or older who report consuming vegetables less than once per day.

Denominator: Number of adults aged 18 years or older who reported information about vegetable consumption (excluding those who refused to answer, had a missing answer, or answered "don't know/not sure").

Measures of frequency: Annual prevalence: crude and age-adjusted (standardized by the direct method to the year 2000 standard U.S. population, distribution 9)[3] with 95% confidence intervals and by demographic characteristics when feasible.

Time period of case definition: Per day.

Ever smoking

Demographic group: Adults aged 18 years or older.

Numerator: Adults aged 18 years or older who report they ever smoked 100 cigarettes.

Denominator: Number of adults aged 18 years or older who reported information about cigarette smoking (excluding those who refused to answer, had a missing answer, or answered "don't know/not sure").

Measures of frequency: Annual prevalence: crude and age-adjusted (standardized by the direct method to the year 2000 standard U.S. population, distribution 9)[3] with 95% confidence intervals and by demographic characteristics when feasible.

Time period of case definition: Lifetime.

Never smoking

Demographic group: Adults aged 18 years or older.

Numerator: Adults aged 18 years or older who report smoking fewer than 100 cigarettes.

Denominator: Number of adults aged 18 years or older who reported information about cigarette smoking (excluding those who refused to answer, had a missing answer, or answered "don't know/not sure").

Measures of frequency: Annual prevalence: crude and age-adjusted (standardized by the direct method to the year 2000 standard U.S. population, distribution 9)[3] with 95% confidence intervals and by demographic characteristics when feasible.

Time period of case definition: Lifetime.

Human papillomavirus (HPV) vaccination coverage

Vaccination against HPV is recommended to prevent new HPV infections and HPV-associated diseases, including some cancers.[5]

Teen vaccination coverage data displayed in U.S. Cancer Statistics Data Visualizations tool are collected through the National Immunization Survey-Teen (NIS-Teen). The NIS-Teen is a random-digit-dialed survey of parents or guardians of teens who are 13 to 17 years old. It has a sample size of more than 20,000 teens. The telephone survey is followed by a questionnaire mailed to vaccination providers to obtain the teen's vaccination history. Vaccination coverage estimates are based on provider-reported vaccination histories. Complex statistical methods are used to adjust for teens whose parents did not participate in the survey, who lived in households without telephones, or whose vaccination histories were not reported by their providers.

The Data Visualizations tool displays HPV vaccination coverage estimates (percentage) by sex for the entire United States, each state, and the District of Columbia. These estimates are for teens aged 13 to 17 years who were reported being up-to-date on HPV vaccination as recommended by the Advisory Committee on Immunization Practices.

References

1. Pickens CM, Pierannunzi C, Garvin W, Town M. Surveillance for certain health behaviors and conditions among states and selected local areas—Behavioral Risk Factor Surveillance System, United States, 2015. *MMWR Surveill Summ.* 2018;67(SS-9):1–90.
2. Holt JB, Huston SL, Heidari K, et al. Indicators for chronic disease surveillance—United States, 2013. *MMWR Recomm Rep.* 2015:64(RR01):1–15.
3. Klein RJ, Schoenborn CA. Age adjustment using 2000 projected U.S. population. Healthy People 2010 statistical notes, no. 20. Hyattsville, MD: U.S. Department of Health and Human Services, CDC, National Center for Health Statistics; 2001.
4. Zhang X, Holt JB, Yun S, Lu H, Greenlund KJ, Croft JB. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the Behavioral Risk Factor Surveillance System. *Am J Epidemiol.* 2015;182(2):127–137.
5. Meites E, Szilagyi PG, Chesson HW, Unger ER, Romero JR, Markowitz LE. Human papillomavirus vaccination for adults: updated recommendations of the Advisory Committee on Immunization Practices. *MMWR Morb Mortal Wkly Rep.* 2019;68:698–702.

Risk Factor-Associated Cancers

Although cancer represents many heterogeneous diseases, some cancer types share common risk factors.[1] For example, conclusive evidence links several cancers with alcohol use, human papillomavirus (HPV) infection, obesity, physical inactivity, and tobacco use.[2 3 4 5 6]

Using standard definitions for risk factor-associated cancers facilitates comparisons of cancer burden across states and communities. Although cancer may occur among people who were not exposed to a risk factor, population-based risk factor-associated cancer rates can help identify communities with high cancer rates. Clinical preventive services and community-based approaches can help reduce risk factors. Cancer surveillance data can track the effectiveness of these approaches.

Definitions of risk factor groupings

Alcohol-associated cancers[2 7]

- Colon and rectum.
- Esophagus.
- Female breast.
- Larynx.
- Liver.
- Oral cavity and pharynx.

HPV-associated cancers[3 8 9 10 11]

- Microscopically confirmed carcinoma of the cervix.
- Squamous cell carcinomas of the vagina, vulva, penis, anus, rectum, and oropharynx.

Obesity-associated cancers[4 5 12]

- Adenocarcinoma of the esophagus.
- Breast (in postmenopausal women).
- Colon and rectum.
- Endometrium (corpus uterus).
- Gallbladder.
- Gastric cardia.
- Kidney (renal cell).
- Liver.
- Ovary.
- Pancreas.
- Thyroid.
- Meningioma.
- Multiple myeloma.

Physical inactivity-associated cancers[5 12]

- Breast cancer (in postmenopausal women).
- Colon.
- Endometrium (corpus uterus).

Tobacco-associated cancers[6]

- Acute myeloid leukemia.
- Cervix.
- Colon and rectum.
- Esophagus.
- Kidney and renal pelvis.
- Larynx.
- Liver.
- Lung, bronchus, and trachea.
- Oral cavity and pharynx.
- Pancreas.
- Stomach.
- Urinary bladder.

The ICD-O-3 site and histology codes used to define these five variables are available in Definitions of Risk Factor-Associated Cancers.

References

1. Henley SJ, Singh SD, King J, Wilson RJ, O'Neil ME, Ryerson AB. Invasive cancer incidence and survival—United States, 2013. *MMWR Morb Mortal Wkly Rep.* 2017;66:69–75.

2. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans. Volume 100E: Personal Habits and Indoor Combustions: Consumption of Alcoholic Beverages. Lyon, France: International Agency for Research on Cancer; 2012.

3. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans. Volume 90: Human Papillomaviruses. Lyon, France: International Agency for Research on Cancer; 2007.

4. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer—viewpoint of the IARC Working Group. *N Engl J Med.* 2016;375:794-798.

5. World Cancer Research Fund / American Institute for Cancer Research. *Diet, Nutrition, Physical Activity and Cancer: A Global Perspective.* Continuous Update Project Expert Report 2018.

6. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.

7. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans: Volume 96: alcohol consumption and ethyl carbamate. Lyon, France: International Agency for Research on Cancer 2010;96.

8. Watson M, Saraiya M, Ahmed F, et al. Using population-based cancer registry data to assess the burden of human papillomavirus-associated cancers in the United States: overview of methods. *Cancer.* 2008;113(10 Suppl):2841–2854.

9. Saraiya M, Unger ER, Thompson TD, et al. U.S. assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. *J Natl Cancer Inst.* 2015;107(6):djv086.

10. Viens LJ, Henley SJ, Watson M, et al. Human papillomavirus–associated cancers—United States, 2008–2012. *MMWR Morb Mortal Wkly Rep.* 2016;65(26):661–666.

11. Centers for Disease Control and Prevention. Cancers linked with HPV each year. Atlanta, GA: U.S. Department of Health and Human Services. Accessed May 16, 2024.

12. Eheman C, Henley SJ, Ballard-Barbash R, et al. Annual report to the nation on the status of cancer, 1975–2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer.* 2012;118:2338–2366.

Incidence and Death Estimates by Congressional District

Cancer death rates and counts for 2018–2022 were estimated for the 436 federal congressional districts according to the boundaries for the 116th Congress of the United States.[1,2]

Cancer incidence rates and counts were estimated for 432 federal congressional districts because county-level incidence data were not available for Kansas (4 congressional districts) due to state legislation and regulations which prohibit the release of county-level data to outside entities.

Data for 27 congressional districts are not presented:

- Illinois (18 congressional districts) opted not to present congressional district-specific estimated case counts and incidence rates.
- Indiana (9 congressional districts) did not meet the publication criteria for diagnosis years 2020 and 2021 and was excluded from the analysis.

Estimated incidence rates and counts are presented from 405 congressional districts.

Methods for creating congressional district estimates

A brief description of the methods for estimating congressional district rates and counts is provided below. For specific inquiries, please email the U.S. Cancer Statistics team at uscsdata@cdc.gov.

Eight congressional districts follow state or federal district boundaries: Alaska, Delaware, District of Columbia, Montana, North Dakota, South Dakota, Vermont, and Wyoming. Those districts were estimated according to the state rates and counts. Rates and counts were calculated for the remaining districts as described below.

Rate calculations

Rates were estimated by assigning the county-level age-adjusted rates to the census block and weighting those by the block population proportion of the congressional district. Those weighted rates were then aggregated over the blocks within the congressional district to estimate the district rate. The rates were age-adjusted to the 2000 US standard population using 15 age groups: 0 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 or older.

More specifically, the following steps were taken:

- The 2010 US Census Summary File 1 was used to determine population estimates within each census block by race and sex. The populations were assigned to congressional districts.
- The age-adjusted county-level rates by race and sex were calculated using SEER*Stat. These rates were merged with the block-level population estimates by county.
- The county rates assigned to the census blocks were weighted by the proportion of the block population within the congressional district and then aggregated over the blocks within the congressional district.

Count calculations

The county counts were weighted by the proportion of the county population in the congressional district to the overall county population. The weighted counts were then aggregated over the counties in the congressional district. This gives the same estimates as weighting at the block level similar to the rate calculations, but is more efficient in terms of computer time. Estimates for both sexes combined were obtained by summing the male estimate and female estimate.

Available data

Estimates are presented by sex (both sexes, male, and female) and race and ethnicity (all races, non-Hispanic White, Black, and Hispanic). Block-level population data were not available by ethnicity for races other than White. As a result, the estimates for Black people include both Hispanic and non-Hispanic Black people. Data are presented for all cancers combined and 20 leading cancers. Data are suppressed for cells with fewer than 16 estimated cases. Data for specific race groups may be suppressed at the state's request.

Since the congressional district estimates require county-level data, if any county data are missing, then the overall state counts presented in the Congressional Districts section will not match the counts in the U.S. Cancer Statistics Data Visualizations tool's State section. Instead, the counts in the Congressional Districts section will match the state counts calculated by aggregating across the U.S. Cancer Statistics county-level data.

References

1. Hao Y, Ward EM, Jemal A, Pickle LW, Thun MJ. U.S. congressional district cancer death rates. *Int J Health Geogr*. 2006;5:28.

2. Siegel RL, Sahar L, Portier KM, Ward EM, Jemal A. Cancer death rates in U.S. congressional districts. *CA Cancer J Clin*. 2015;65(5):339–344.

Suppression of Rates and Counts

Suppression for reliability

When the numbers of cases or deaths used to compute rates are small, those rates tend to have poor reliability.[1] Therefore, to discourage misinterpretation or misuse of rates or counts that are unstable, incidence and death rates and counts are not shown in tables and figures when the case or death counts are below 16.

A count of fewer than 16 results in a standard error of the rate that is approximately 25% (or more) larger than the rate itself. Similarly, a case count below 16 results in the rate's 95% confidence interval width being at least as large as the rate itself. These relationships were derived under the assumption of a Poisson process and with the standard population age distribution assumed to be similar to the observed population age distribution. A suppressed rate does not necessarily mean that the rate was low.

Suppression for confidentiality

Another important reason for using a threshold value for suppressing cells is to protect the confidentiality of patients whose data are included in a report by reducing the risk of disclosing their identity.[2] The cell suppression threshold value of 16 is more than sufficient to protect patient confidentiality.

Suppression for other reasons

While data meet the U.S. Cancer Statistics publication criteria, a central cancer registry may suppress its data for various reasons. For example, a registry's algorithms to correct for unknown race or ethnicity may not function properly. In these circumstances, data are suppressed upon the state's request.

State and county data

Cancer incidence rates are presented for each county or county equivalent as available for the most recent 5-year period. County data are not available from Kansas because state legislation and regulations prohibit the release of county-level data to outside entities. Additional data suppressions are applied in accordance with the rules outlined above.

County-level mortality data are not presented for Connecticut because population estimates for the former 8 legacy counties are not available for 2022. Connecticut population estimates for 2022 are reported for 9 planning regions as county-equivalent areas, instead of the former 8 legacy counties in the Vintage 2022 postcensal series released by the Census Bureau on June 22, 2023.

Total United States

Cancer incidence rates for the United States are aggregate rates based on cancer cases reported from central cancer registries that met the U.S. Cancer Statistics publication criteria. They are the best estimates of the U.S. cancer burden available that are based on observed data. Case counts for the U.S. incidence rates for all ages combined are presented. Puerto Rico data are not included in the total U.S. counts and corresponding rate calculations.

References

1. Brillinger DR. The natural variability of vital rates and associated statistics. *Biometrics.* 1986;42(4):693–734.

2. Doyle P, Lane JI, Theeuwes JM, Zayatz LM. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies.* Amsterdam, The Netherlands: Elsevier Science; 2001.

Interpreting Incidence Data

Each year when U.S. Cancer Statistics data are released, we update data products with the most recent data submission. Users of cancer incidence data published by federal agencies should be mindful of the data submission dates for all data used in their analyses.

Choice of standard population and population denominator

The U.S. Department of Health and Human Services' policy for reporting death and disease rates was motivated by a need to standardize age adjustment procedures across government agencies.[1] [2] Because of the aging U.S. population, the 2000 US standard population gives more weight to older age categories than the 1940 and 1970 standard populations did.[2] The National Center for Health Statistics (NCHS) regularly evaluates the population standard and currently recommends using the 2000 U.S. standard population for calculating age-adjusted rates.

The age-adjusted rates in the Data Visualizations tool should not be compared with cancer incidence rates adjusted to different standard populations.

Incidence rates also are influenced by the choice of population denominators used in calculating these rates. Because some state health departments use customized projections of the state's population when calculating incidence rates, the rates in the Data Visualizations tool may differ slightly from those published by individual states.

Registries' data quality

Data quality is evaluated routinely by CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) Program.[3] [4] Comprehensive evaluation activities are conducted to find missing cases or to identify errors in the data. Although the cancer registries meet data quality criteria for all invasive sites combined, the completeness and quality of site-specific data may vary, making in-depth analyses critical to present reliable data.

The observed rates may have been influenced by differences in the timeliness, completeness, and accuracy of the data from one registry to another, from one reporting period to another, or from one cancer site to another. In rare instances, a registry may identify a data quality issue after the file is submitted to CDC. In those instances, CDC will either suppress the identified segment or exclude the registry's data from analytic products.

Reporting

Reporting time intervals

Completeness and accuracy of the site-specific data also may be affected by the time interval allowed for reporting data to the two federal programs. The NPCR and SEER time interval for reporting data differ: For each submission year, NPCR allows a 23-month interval after the close of the diagnosis year and SEER allows a 22-month interval.

Reporting delays

Delays in reporting cancer cases can affect timely and accurate calculation of cancer incidence rates.[5] Cases are reported continuously to state and metropolitan area cancer registries in accordance with statutory and contractual requirements.

After the initial submission of the most recent year's data to the federal funding agency, cancer registries update their data on the basis of new information received. Therefore, some cancer cases likely will have been reported to state and metropolitan-area cancer registries after the registries submitted their data to CDC or NCI. For this reason, incidence rates and case counts reported directly by state or metropolitan area cancer registries may differ from those that appear in the Data Visualizations tool.

Reporting delays appear to be more common for cancers that usually are diagnosed and treated in non-hospital settings such as physicians' offices (for example, early-stage prostate and breast cancers and melanoma of the skin). Efforts are underway to reduce reporting delays. Methods to adjust incidence rates for reporting delay were not applied to the data in this report.[5]

Continual data updates

Each year, central cancer registries submit an updated version of previous years' data and data for a new diagnosis year to CDC, NCI, or both agencies. Federal agencies, in turn, update their cancer incidence statistics with each data submission and document the registries' date of data submission whenever the data are published. These continual updates illustrate the dynamic nature of cancer surveillance and the attention to detail that characterizes cancer registries. Each year when U.S. Cancer Statistics data are released, we update data products with the most recent data submission.

Geographic variation

Geographic variation in cancer incidence rates may result from regional differences in the exposure of the population to known or unknown risk factors.[6 7 8 9] Differences may arise because of differences in:

- Sociodemographic characteristics of the population (age, race and ethnicity, geographic region, urban or rural residence).
- Screening test use.
- Health-related behaviors such as tobacco use, diet, and physical activity.
- Exposure to cancer-causing agents.
- Factors associated with the registries' operations, such as completeness, timeliness, and specificity in coding cancer sites.

Researchers are investigating variability associated with known factors that may affect cancer rates. Researchers use model-based statistical techniques and other approaches for surveillance research. Differences in registry operations are being evaluated to ensure consistency and quality in reporting data.

References

1. Anderson RN, Rosenberg HM. Report of the Second Workshop on Age Adjustment.*Vital Health Stat 4*. 1998;(30):I–VI, 1–37.

2. Anderson RN, Rosenberg HM. Age standardization of death rates: implementation of the year 2000 standard. *Natl Vital Stat Rep.* 1998;47(3):1–16, 20.

3. Fritz A. The SEER Program's commitment to data quality. *J Registry Manag.* 2001;28(1):35–40.

4. Hutton MD, Simpson LD, Miller DS, Weir HK, McDavid K, Hall HI. Progress toward nationwide cancer surveillance: an evaluation of the National Program of Cancer Registries, 1994–1999. *J Registry Manag.* 2001;28(3):113–120.

5. Clegg LX, Feuer EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. *J Natl Cancer Inst.* 2002;94(20):1537–1545.

6. Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System Operational and User's Guide. Version 3.0.* Atlanta (GA): Centers for Disease Control and Prevention; 2005.

7. Devesa SS, Grauman DJ, Blot WJ, Pennello GA, Hoover RN. *Atlas of Cancer Mortality in the United States, 1950–1994.* Bethesda (MD): National Cancer Institute; 1999.

8. Howe HL, Keller JE, Lehnherr M. Relation between population density and cancer incidence, Illinois, 1986–1990. *Am J Epidemiol.* 1993;138(1):29–36.

9. Wingo PA, Jamison PM, Hiatt RA, et al. Building the infrastructure for nationwide cancer surveillance and control—a comparison between the National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program (United States). *Cancer Causes Control.* 2003;14(2):175–193.

Interpreting Mortality Data

Cause of death determined by autopsy combined with clinical data is considered the best estimate of the true cause of death.[1]

Autopsy studies of mortality data coded according to the eighth or ninth revision of the *International Classification of Diseases* (ICD) (ICD-8A or ICD-9) indicate that, when neoplasms (cancers) are an underlying cause of death, the sensitivity of death certificates was 87% to 93%, and their positive predictive value was 85% to 96%.[1 2 3] However, these studies are limited by selection bias, and fewer than 10% of deaths in the United States are autopsied.[4]

Death certificates' reliability

The percentage of cancers coded as the underlying cause of death on the death certificate that agree with the cancer diagnosis in the medical record is an indication of the reliability with which the underlying cause of death can be determined from the death certificate.

In a study by German et al., central cancer registry records from California, Colorado, and Idaho were linked with state vital statistics data and evaluated by demographic and tumor information across 79 site categories. A retrospective arm (confirmation rate per 100 deaths) compared death certificate data from 2002 to 2004 with cancer registry diagnoses from 1993 to 2004, while a prospective arm (detection rate per 100 deaths) compared cancer registry diagnoses from 1993 to 1995 with death certificate data from 1993 to 2004 by International Statistical Classification of Diseases and Related Health Problems (ICD) version used to code deaths.

The overall confirmation rate for ICD-10 was 82.8% (95% confidence interval [CI], 82.6–83.0%), the overall detection rate for ICD-10 was 81.0% (95% CI, 80.4–81.6%), and the overall detection rate for ICD-9 was 85.0% (95% CI, 84.8–85.2%). These rates varied across primary sites, where some rates were less than 50%, some were 95% or greater, and notable differences between confirmation and detection rates were observed. Confirmation or detection rates were 95% or greater for some of the most common cancers in the United States, such as prostate, breast, and lung and bronchus. This study recorded important unique information on the quality of cancer mortality data obtained from death certificates, particularly underlying causes of death coded in ICD-10.[5]

Improving the accuracy of vital statistics

CDC's National Center for Health Statistics has worked with the Social Security Administration and the National Association for Public Health Statistics and Information Systems to develop and promote electronic systems to improve the accuracy and timeliness of vital statistics. Standard certificates for births and deaths were revised, and state vital registration systems are being updated to collect data electronically. These systems will accommodate certificate revisions, special studies or projects, and linkage with other health promotion programs. Guides are routinely updated and provided to professionals who complete death certificates.

References

1. Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. *N Engl J Med.* 1985;313(20):1263–1269.

2. Engel LW, Strauchen JA, Chiazze L Jr, Heid M. Accuracy of death certification in an autopsied population with specific attention to malignant neoplasms and vascular diseases. *Am J Epidemiol.* 1980;111(1):99–112.

3. Schottenfeld D, Eaton M, Sommers SC, Alonso DR, Wilkinson C. The autopsy as a measure of accuracy of the death certificate. *Bull N Y Acad Med.* 1982;58(9):778–794.

4. Sinard JH. Factors affecting autopsy rates, autopsy request rates, and autopsy findings at a large academic medical center. *Exp Mol Pathol.* 2001;70(3):333–343.

5. German RR, Fink AK, Heron M, et al. The accuracy of cancer mortality statistics based on death certificates in the United States. *Cancer Epidemiol.* 2011;35(2):126–131.

Interpreting Race and Ethnicity in Cancer Data

The North American Association of Central Cancer Registries (NAACCR) Race and Ethnicity Identifier Assessment Project confirmed the importance of publishing cancer rates by race and ethnicity (specifically, Hispanic origin).[1]

Cancer incidence

When reporting cancer incidence, race and ethnicity information is abstracted from medical records and grouped into race and ethnicity categories.[2] Although registries use standardized data items and codes for both race and ethnicity, the initial collection of this information by health care facilities and practitioners and the procedures for assigning and verifying codes for race and ethnicity are not well standardized.[1] Thus, some inconsistency is expected in this information.

Cancer mortality

When reporting cancer mortality, race and Hispanic origin are recorded separately on the death certificate by the funeral director as provided by an informant or on the basis of observation.[3] Inconsistencies in the collection and coding of data on race and Hispanic origin and their effect on mortality statistics have been described.[4]

Effects of misclassification

The net effect of misclassification is greatest for American Indian and Alaska Native people. Misclassification is smaller for Asian and Pacific Islander people and Hispanic people, and minimal for Black people and White people. Therefore, incidence and mortality data published in this report may be underestimated for Asian and Pacific Islander, American Indian and Alaska Native, and Hispanic people. This may be due to racial and Hispanic origin misclassification.

Improving data accuracy

In the U.S. Cancer Statistics Data Visualizations tool, we have restricted analysis to non-Hispanic populations to overcome racial misclassification and to represent populations more accurately. CDC's National Center for Health Statistics is working with states to improve the reporting of race and ethnicity on death certificates.

The Data Visualizations tool presents cancer incidence data for all races combined and by bridged race and ethnicity (non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, non-Hispanic Asian and Pacific Islander, and Hispanic) categories. Starting with 2018 deaths, mortality data are presented for all races combined and by single-race race and ethnicity (non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, non-Hispanic Asian, non-Hispanic Native Hawaiian or Other Pacific Islander, and Hispanic. Use caution while comparing with data in Archived Reports as the differences in rates are due to changes in data presentation and not due to rate changes. Also use caution if comparing bridged race and ethnicity incidence data to single race and ethnicity mortality data as single race and bridged race data are not considered to be directly comparable.[5] Puerto Rico data are available only for all races and ethnicities combined.

Asian and Pacific Islander people

Central cancer registries have codes for race that allow them to document the occurrence of cancer in 25 Asian and Pacific Islander subpopulations.[2] But the subpopulations are grouped into a single Asian and Pacific Islander category because of small numbers and concerns about possible misclassification. The Asian and Pacific Islander category is restricted to non-Hispanic persons.

Studies show excellent agreement (k=0.90) between Asian and Pacific Islander race in Surveillance, Epidemiology, and End Results (SEER) registry data and self-reported data from the U.S. Census.[6] Studies examined the misclassification of race for Asian and Pacific Islander subpopulations in cancer registries.[6 7 8 9] Nearly all National Program of Cancer Registries (NPCR) and SEER registries assigned Asian, not otherwise specified to a more specific Asian race through the standardized use of the NAACCR Asian and Pacific Islander Identification Algorithm (NAPIIA) version 1.2.

For cases reported in 1999, Kansas opted not to present state- and county-specific counts and rates for non-Hispanic Asian and Pacific Islander persons. The national rates presented include data for Kansas.

A study reported 90% agreement between Asian and Pacific Islander race reported on death certificates and self-reported data from the U.S. Census.[4]

Hispanic people

The overall agreement between Hispanic ethnicity collected by SEER registries and self-reported ethnicity from the U.S. Census was substantial (k=0.61). Hispanic people were found to be underclassified in the SEER data compared to self-reports.[6] Nearly all NPCR and SEER registries assigned Hispanic ethnicity through the standardized use of the NAACCR Hispanic Identification Algorithm (NHIA) version 2 (NHIA v2).[10] After applying the NHIA v2, cases not classified as Hispanic are classified as non-Hispanic, leaving no cases with unknown Hispanic status.

A study reported an 88% record-by-record agreement between Hispanic origin on death certificates and self-reported data.[4]

Death counts and rates for Hispanic people are presented at the national and state levels for all 50 states and the District of Columbia. Hispanic origin is assigned to cancer mortality data on the basis of information collected from death certificates.

Improving estimation of cancer burden among American Indian and Alaska Native people

More American Indian and Alaska Native patients are misclassified as another race in cancer registry records than patients in other racial groups. Studies have found that this racial misclassification contributes to underestimates of cancer incidence and death rates among the American Indian and Alaska Native population.[4 11] Accurate determination of disease burden is a critical first step toward identifying health disparities. Methods that can improve the accuracy of cancer burden estimates among the American Indian and Alaska Native population are described below.

Linkage with Indian Health Service administrative records

The Indian Health Service (IHS) provides medical services to American Indian and Alaska Native people who are enrolled members of federally recognized tribes. The IHS provides health care to about 2.2 million people, a number equivalent to about 64% of the U.S. American Indian and Alaska Native population.[11] While IHS coverage of these populations varies by region, it does not include American Indian and Alaska Native people who are members of non-federally recognized tribes, and underrepresents those who live in certain urban areas. People who are eligible to receive IHS services have sufficient native ancestry in a federally recognized tribe to be classified accurately as an American Indian or Alaska Native person.

As a standard practice, central cancer registries classify race as coded in the medical record. To address American Indian and Alaska Native misclassification in cancer registry data, selected registries in CDC's NPCR and all registries in the National Cancer Institute's SEER program linked their central cancer registry data to the IHS administrative records database for cases diagnosed from 1995 to 2021 and 1988 to 2021, respectively. Results of the linkage were captured in the data element, IHS Link (NAACCR data item 192).[2] Central cancer registries include race and IHS Link in their annual data submissions to CDC or NCI. Using the race and IHS Link data elements, CDC and NCI created a recoded race variable. If a cancer case had an IHS Link value that indicated a match to IHS and race is White, other, or unknown, then the recoded race variable was coded as American Indian and Alaska Native.

Although the linkage with IHS does not completely resolve the classification of race for American Indian and Alaska Native cases, it helps provide a more comprehensive and accurate picture of the cancer burden in this population.

Restriction to non-Hispanic populations

Updated bridged intercensal population estimates significantly overestimated the number of American Indian and Alaska Native persons of Hispanic origin.[11 12 13] Because these population estimates are used as denominators in rate calculations, larger than expected denominators can result in underestimation of rates. Studies demonstrate that restricting analysis to non-Hispanic populations can improve the accuracy of cancer incidence and death rate estimates among American Indian and Alaska Native people.[13]

Restriction to IHS Purchased/Referred Care Delivery Areas

The IHS Purchased/Referred Care Delivery Area (PRCDA) is the geographic area within which the IHS makes purchased or referred care available to members of an identified Indian community who reside in the area. The IHS uses it to determine eligibility for services not directly available within the IHS.

The IHS PRCDA consists of counties that include all or part of an American Indian or Alaska Native reservation or have a common boundary with a federally recognized tribal land, as defined in the October 14, 2020 *Federal Register* (82 FR 47004). There are 36 states that have at least one PRCDA-designated county. The PRCDA counties have higher proportions of American Indian and Alaska Native people in relation to the total population than non-PRCDA counties, with 53.5% of the U.S. American Indian and Alaska Native population residing in the 685 counties designated as PRCDA. Linkage studies have indicated more accurate race classification for American Indian and Alaska Native persons in PRCDA counties.[11 12 13 14 15 16 17]

Data on American Indian and Alaska Native people in the U.S. Cancer Statistics Data Visualizations tool

The U.S. Cancer Statistics Data Visualizations tool presents national, state, and county data by race, including non-Hispanic American Indian and Alaska Native people. The national data include non-Hispanic American

Indian and Alaska Native populations in all U.S. counties. These data use the results from the linkage with IHS to classify race, and are restricted to non-Hispanic people only. As described in the section above, these restrictions can improve the accuracy of cancer burden estimates among the American Indian and Alaska Native population.

State- and county-specific data for non-Hispanic American Indian and Alaska Native persons are not presented for states that opted not to present these data: Illinois, Kansas, New Jersey, and New York.

Data on American Indian and Alaska Native people in the At a Glance section

The U.S. Cancer Statistics Data Visualizations tool's American Indian and Alaska Native restricted to PRCDA only module presents data from the United States Cancer Statistics American Indian and Alaska Native Incidence Analytic Database (USCS AIAD) in the tool's At a Glance section. This database uses the three methods described above to improve the accuracy of cancer burden estimates among American Indian and Alaska Native people:

- First, this database uses the recoded race variable to classify race. Only people of American Indian and Alaska Native race or White race (as comparison) are included in the module.
- Second, the database is restricted to persons of non-Hispanic origin.
- Third, the database is restricted to persons residing in PRCDA counties.

This database includes data elements specific to the American Indian and Alaska Native population, such as IHS Region and PRCDA county.

The USCS AIAD data can be displayed for all IHS regions combined or by six IHS regions: Alaska, Pacific Coast, Southwest, Northern Plains, Southern Plains, and East. The states grouped by IHS region are:

- Alaska: Alaska.
- Pacific Coast: California, Idaho, Oregon, and Washington.
- Southwest: Arizona, Colorado, Nevada, New Mexico, and Utah.
- Northern Plains: Indiana, Iowa, Michigan, Minnesota, Montana, Nebraska, North Dakota, South Dakota, Wisconsin, and Wyoming.
- Southern Plains: Kansas, Oklahoma, and Texas.
- East: Alabama, Connecticut, Florida, Louisiana, Massachusetts, Maine, Mississippi, New York, North Carolina, Pennsylvania, Rhode Island, South Carolina, and Virginia.

The percentages of the American Indian and Alaska Native population living in PRCDA-designated counties by IHS region from 2017 to 2021 were:

- Alaska: 100%.
- Pacific Coast: 61.0%.
- Southwest: 86.2%.
- Northern Plains: 53.9%.
- Southern Plains: 55.7%.
- East: 17.6%.
- Total United States: 53.5%.

Studies have shown substantial variation in rates in the American Indian and Alaska Native population by IHS region.[18 19] IHS regions have been presented in several publications focusing on American Indian and Alaska Native people. This approach was determined to be preferable to the use of smaller jurisdictions, such as IHS Administrative Areas, which yielded less stable estimates.[14 20 21 22]

References

1. O'Malley C, Hu KU, West DW. *North American Association of Central Cancer Registries: Race and Ethnicity Identifier Assessment Project.* Springfield (IL): North American Association of Central Cancer Registries; 2001.

2. Havener L, Hulstrom D. *Standards for Cancer Registries Vol. II: Data Standards and Data Dictionary.* 10th ed., version 11. Springfield (IL): North American Association of Central Cancer Registries; 2004.

3. Miniño AM, Heron MP, Smith BL, Kochanek K. Deaths: final data for 2004. *Natl Vital Stat Rep.* 2007;55(19).

4. Arias E, Heron M, Hakes JK. The validity of race and Hispanic-origin reporting on death certificates in the United States: an update. *Vital Health Stat 2.* 2016;2(172).

5. Heron MP. Comparability of race-specific mortality data based on 1977 versus 1997 reporting standards. *Natl Vital Stat Rep*. 2021;70(3):1–31.

6. Clegg LX, Reichman ME, Hankey BF, et al. Quality of race, Hispanic ethnicity, and immigrant status in population-based cancer registry data: implications for health disparity studies. *Cancer Causes Control.* 2007;18(2):177–187.

7. NAACCR Race and Ethnicity Work Group. *NAACCR Asian Pacific Islander Identification Algorithm [NAPIIA v1.2.1].* Springfield (IL): North American Association of Central Cancer Registries; 2008.

8. Boscoe FP. Issues with the coding of Asian race in cancer registration. *J Registry Manag.* 2007;34(4):135–139.

9. Boscoe FP, Schymura MJ, Hsieh M, Williams MA, Henry KA. Issues with the coding of Pacific Islanders in central cancer registries. *J Registry Manag.* 2008;35(2):47–51.

10. NAACCR Race and Ethnicity Work Group. NAACCR Guideline for Enhancing Hispanic/Latino Identification: Revised NAACCR Hispanic/Latino Identification Algorithm [NHIA v2.2.1]. Springfield (IL): North American Association of Central Cancer Registries. September 2011.

11. Jim MA, Arias E, Seneca DS, et al. Racial misclassification of American Indians and Alaska Natives by Indian Health Service Contract Health Service Delivery Area. *Am J Public Health.* 2014;104(6 suppl 3):S29–S302.

12. Arias E, Xu J, Jim MA. Period life tables for the non-Hispanic American Indian and Alaska Native population, 2007–2009. *Am J Public Health.* 2014;104:S312–S319.

13. Espey DK, Jim MA, Richards TB, Begay C, Haverkamp D, Roberts D. Methods for improving the quality and completeness of mortality data for American Indians and Alaska Natives. *Am J Public Health.* 2014;104:S286–S294.

14. Espey DK, Wiggins CL, Jim MA, Miller BA, Johnson CJ, Becker TM. Methods for improving cancer surveillance data in American Indian and Alaska Native populations. *Cancer.* 2008;113(5 suppl):1120–1130.

15. Sugarman JR, Holliday M, Ross A, Castorina J, Hui Y. Improving American Indian cancer data in the Washington State Cancer Registry using linkages with the Indian Health Service and tribal records. *Cancer.* 1996;78(7 Suppl):1564–1568.

16. Frost F, Taylor V, Fries E. Racial misclassification of Native Americans in a Surveillance, Epidemiology, and End Results cancer registry. *J Natl Cancer Inst.* 1992;84(12):957–962.

17. Kwong SL, Perkins CL, Snipes KP, Wright WF. Improving American Indian cancer data in the California Cancer Registry by linkage with the Indian Health Service. *J Registry Manag.* 1998;25(1):17–20.

18. Wiggins CL, Espey DK, Wingo PA, et al. Cancer among American Indians and Alaska Natives in the United States, 1999–2004. *Cancer.* 2008;113(5 suppl):1142–1152.

19. White MC, Espey DK, Swan J, Wiggins CL, Eheman C, Kaur JS. Disparities in cancer mortality and incidence among American Indians and Alaska Natives in the United States. *Am J Public Health.* 2014;104:S377–S387.

20. Espey DK, Wu XC, Swan J, et al. Annual report to the nation on the status of cancer, 1975–2004. *Cancer.* 2007;110(10):2119–2152.

21. Espey DK, Jim MA, Richards TB, Begay C, Haverkamp D, Roberts D. Methods for improving the quality and completeness of mortality data for American Indians and Alaska Natives. *Am J Public Health.* 2014;104:S286–S294.

22. Melkonian SC, Chen L, Jim MA, Haverkamp D, King JB. Disparities in incidence and trends of colorectal, lung, female breast, and cervical cancers among non-Hispanic American Indian and Alaska Native people, 1999–2018. *Cancer Causes Control.* 2023;34(8):657–670.

Guidance for Comparing Cancer Data by Geographic Region

When looking at figures that rank cancer rates by state, county, or other geographic region, some readers wonder why their community has higher rates than other communities or the national average. For example, some may worry that exposure to environmental carcinogens may be responsible, when several other explanations are more likely. Consider the following points when comparing rates by geographic region.

Difference in cancer rates

Differences among racial and ethnic populations

Some cancers have different rates for different racial and ethnic populations. For example, breast cancer incidence rates are usually higher in non-Hispanic White women than in women of other racial and ethnic populations. Prostate cancer incidence rates are higher in non-Hispanic Black men. Therefore, when comparing cancer rates across geographic regions, consider the racial makeup of the region's population, which is determined through the statistical adjustment of rates by race and ethnicity. Presenting rates for specific racial and ethnic populations may be preferable and is more easily understood by a lay audience.

Variations in risk factors and health behaviors

Some differences in cancer rates among geographic regions may be explained by differences in known risk factors among the populations of those regions. For example, rates of lung cancer and other tobacco-associated cancers are higher in regions with a higher prevalence of smoking. Although environmental carcinogens are responsible for some cancer cases, most cases appear to be related to lifestyle factors such as smoking. Geographic variations in cancer rates are thought largely to reflect variations in these lifestyle factors.[1,2] Additionally, many people move during their lifetime and may have been exposed to a risk factor in one community and diagnosed with cancer years later in another community.

Variations in medical care

Variations among geographic regions in medical care factors may also result in differences in cancer rates. In places where higher percentages of the population participate in cancer screening, more cancers will be diagnosed. Screening leads to earlier detection of tumors that have a better prognosis and may find tumors that grow slowly. Therefore, the cancer incidence rate only tells part of the story.

Influence of aging on cancer rates

The likelihood of being diagnosed with cancer increases with age. These rates have been adjusted for age so that geographic regions can be compared without concern that differences in their rates result from differences in the age distribution of their populations. However, this adjustment may be imperfect if the relationship between age and cancer risk is not the same for all regions.

Additional considerations

Measuring burden

The importance of cancer as a public health problem in a community is more a function of the absolute rate of cancer rather than the community's relative ranking in incidence or mortality. For example, Utah has proportionately fewer people who have ever smoked cigarettes than other states, and also has the lowest lung cancer incidence rate of any state. Nevertheless, in Utah lung cancer kills more people than any other cancer—a fact that might be overlooked if one focused only on its low incidence ranking compared with other states.

The true burden of cancer on a community's health care system and economy is determined by the number of people diagnosed with and dying of cancer—not by the age-adjusted cancer rate. So, the cancer rate in one community may seem high compared with other communities, but the number of cases is small.

Accuracy and completeness in cancer reporting and coding

U.S. Cancer Statistics incidence data are from central cancer registries that have high-quality data. However, some information about the patient or tumor may be missing or incomplete. For example, a medical record may only have a PO Box for a patient's address, so the patient is coded as living in that county when they may actually live in a different county. Also, reporting facilities may fall behind on reporting their cancer cases, and registries have varying resources to find these unreported cases. When comparing rates, especially from relatively small geographic regions, remember that each rate depends on the accurate and complete ascertainment of many details.

Random factors and cancer rates

There is some uncertainty in computed cancer rates because many factors contribute to the rates, and some factors may happen at random. Chance plays a role in determining if and when cancer develops in an individual, if the cancer is detected, if the information is entered into the cancer registry, and if the cancer leads to death. For these reasons, rates are expected to vary from year to year within a community even in the absence of a general trend. Caution is warranted when examining cancer rates for a single year, especially when the rates are based on a relatively small number of cases.

Confidence intervals

A 95% confidence interval for the rate is an interval that is expected to contain the true underlying rate 95% of the time. Confidence intervals around the observed age-adjusted rates are available to help interpret the results. Because of the variation in the population sizes and number of reported cases and deaths across geographic regions, there is more uncertainty in the incidence and death rates for some regions compared with others.

The confidence intervals provide a measure of the variability in the rates and some perspective for making comparisons. However, using overlapping confidence intervals to conclude that rates are not significantly different is not recommended. This is a conservative method because it may fail to detect significant differences more often than standard statistical hypothesis testing.

Public health importance

Another consideration when comparing differences between rates is their public health importance. For some rates, numerators and denominators are large and the standard errors are therefore small. This results in statistically significant differences that may be too small to be important for decisions related to population-based public health programs.

References

1. US Burden of Disease Collaborators; Mokdad AH, Ballestros K, et al. The state of U.S. health, 1990–2016: burden of diseases, injuries, and risk factors among U.S. states. *JAMA.* 2018;319(14):1444–1472.
2. Anand P, Kunnumakkara AB, Sundaram C, et al. Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res.* 2008;25(9):2097–2116.