U.S. Cancer Statistics Data Visualizations Tool Technical Notes

November, 2017 Submission
Diagnosis Years 1999–2015

Introduction

The Impact of Cancer

Cancer is the second-leading cause of death among Americans. One of every four deaths in the United States is due to cancer.[1] [2] The 2018 release of United States Cancer Statistics data indicates in 2015 (the most recent year of incidence* data available), 1,633,390 Americans received a new diagnosis of invasive cancer† and 595,919 Americans died from this disease.[3] These counts do not include *in situ* cancers or the more than 1 million cases of basal and squamous cell skin cancers. The National Cancer Institute estimated that on January 1, 2015, 15.1 million Americans were alive with a history of invasive cancer.[4]

The Agency for Healthcare Research and Quality (AHRQ) estimated that for 2014, the direct medical costs‡ for cancer, including all health care expenditures, were $87.3 billion, of which 58.1% was spent on hospital outpatient or office-based provider visits, 27.0% on inpatient hospital stays, and 12.4% on prescription medications.[5]

*2015 is the most recent year for which incidence data are available. These data include cancer deaths during 1999 through 2015. Cancer mortality data for 2017 are available and can be accessed at CDC's National Center for Health Statistics (NCHS) National Vital Statistics System (NVSS).

†Data are from selected central cancer registries, covering 100% of the U.S. population, that meet the data quality criteria for all invasive cancer sites combined. See registry-specific data quality information.

‡The estimates of direct costs are obtained from the AHRQ's Medical Expenditure Panel Survey.

Cancer Prevention

Several effective 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, we must reduce behavioral and environmental factors that increase cancer risk and ensure that high-quality screening services and evidence-based treatments are available and accessible to everyone, including medically underserved populations.[6] [7] The Centers for Disease Control and Prevention's (CDC) Division of Cancer Prevention and Control (DCPC) has supported all 50 states, D.C., tribes and tribal organizations, Pacific Island Jurisdictions, and Puerto Rico in developing comprehensive cancer control plans, which include proven strategies and planned actions to prevent cancer in their geographic regions.[8]

How Cancer Data Are Collected

Cancer registries collect population-based data about 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, and the outcome of treatment and clinical management (survival and vital status).[9] [10] Cancer incidence data are reported to metropolitan area, regional, and statewide cancer registries from a variety of medical facilities, including hospitals, physicians' offices, radiation facilities, freestanding surgical centers, and pathology laboratories. Death data, including deaths due to cancer, are recorded on death certificates that are sent to state 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

Information derived from population-based central cancer registries and from state vital statistics systems is critical for directing effective geographic area and population-specific cancer prevention and control programs that focus on preventing behaviors that put people at increased risk for cancer (such as smoking), and on reducing environmental risk factors (such as occupational exposure to known carcinogens). This information also is essential for deciding which geographic areas should have cancer screening programs, and for making long-term plans for adequate diagnostic and treatment services. Combined data at the national, regional, state, and county levels help federal and state public health officials establish, prioritize, and monitor national initiatives in public health surveillance and track progress toward the national goals and objectives set forth in Healthy People.

Additional resource: Archive of the Annual Reports to the Nation

References

1. Centers for Disease Control and Prevention. *The Burden of Chronic Diseases and Their Risk Factors: National and State Perspectives 2004.* Atlanta (GA): Centers for Disease Control and Prevention; 2005.

2. American Cancer Society. *Cancer Facts and Statistics.* Atlanta (GA): American Cancer Society; accessed June 4, 2018.

3. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on November 2017 submission data (1999–2015): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, June 2018.

4. Noone AM, Howlader N, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975–2015, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER website, April 2018.

5. Agency for Healthcare Research and Quality. Table 3: Total Expenses and Percent Distribution for Selected Conditions by Type of Service: United States, 2014. Medical Expenditure Panel Survey Household Component Data. Generated interactively April 25, 2018.

6. Curry SJ, Byers T, Hewitt M. *Fulfilling the Potential of Cancer Prevention and Control.* Washington (DC): The National Academies Press; 2003.

7. 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.

8. National Comprehensive Cancer Control Program. *Comprehensive Cancer Control Plans.* Atlanta: U.S. Department of Health and Humans Services, Centers for Disease Control and Prevention.

9. American College of Surgeons Commission on Cancer. *Standards of the Commission on Cancer Vol II: Registry Operations and Data Standards (ROADS).* Chicago (IL): American College of Surgeons; 1998.

10. Fritz A, Ries LAG. *The SEER Program Code Manual, third edition.* Bethesda (MD): National Cancer Institute; 1998.

Contributors

Contributor: National Program of Cancer Registries (NPCR)

NPCR registries cover 97% of the U.S. population. In November 2017, CDC received information on more than 28.5 million invasive cancer cases diagnosed from January 1, 1995 through December 31, 2015. More than 1 million new invasive cancer cases are added each year.

Contributor: Surveillance, Epidemiology, and End Results (SEER) Program

The National Cancer Institute's SEER Program collects and publishes data on cancer incidence and survival from 15 population-based cancer registries and 3 supplemental registries covering approximately 30% of the U.S. population.

Contributor: National Vital Statistics System (NVSS)

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

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.

National Program of Cancer Registries (NPCR)

Recognizing the need for more complete local, state, regional, and national data on cancer incidence, in 1992 Congress established the National Program of Cancer Registries (NPCR) by enacting the Cancer Registries Amendment Act, later incorporated into the Public Health Service (PHS) Act [42 U.S.C. 242k]. Congress mandated CDC to provide funds to state and territorial health departments (or their authorized agencies) at a ratio of 3:1 to match state support for the central cancer registry. As of 2018, CDC funds 50 cancer registries: 46 states, the District of Columbia, Puerto Rico, the Pacific Island Jurisdictions, and the U.S. Virgin Islands.

NPCR continues to—

- Monitor the state and national burden of cancer.
- Identify variation in cancer incidence for racial and ethnic populations and for regions within a state, between states, and between regions.
- Provide data for research.
- Provide guidance for the allocation of health resources.
- Respond to public concerns and inquiries about cancer.
- Improve planning for future health care needs.
- Evaluate 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 registries for inclusion in those states' incidence data and are transmitted to CDC as part of the state's annual data submission.

In November 2017, CDC received information on more than 28.5 million invasive cancer cases diagnosed during 1995 through 2015. More than 1 million new invasive cancer cases are added each year.

In conjunction with the annual release of United States Cancer Statistics (USCS) data, CDC's NPCR recognizes each funded central cancer registry for its achievement of the NPCR Standards for Data Completeness, Timeliness, and Quality.

All standards are indicative of complete, timely, and quality data available for cancer control activities addressing the burden on U.S. citizens. Meeting these standards allows inclusion of the program's data in USCS data products.

The release of USCS data in products including the Data Visualizations Tool and Public Use Database exemplifies the progress achieved in creating a national system of cancer surveillance. NPCR commends all who are involved in the collection, analysis, and reporting of cancer incidence and mortality data. Data from regional, 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. USCS data products and many advances in cancer surveillance in the United States would have been impossible without the tireless efforts and many achievements of these organizations.

USCS data products include—

- A web-based data visualization website of USCS data, the official federal cancer statistics.
- Public use databases for researchers to analyze more than 24 million cases of de-identified data reported by NPCR- and SEER-funded sites.
- A public-use data set of pre-calculated cancer incidence rates on CDC WONDER.
- Fact sheets on the states' cancer burden intended for lay audiences.
- 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

In 1971, Congress passed the National Cancer Act, which mandated the collection, analysis, and dissemination of data useful for the prevention, diagnosis, and treatment of cancer. This mandate led to the establishment of the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) Program.[1]

The SEER Program continues to—

- Monitor the burden of cancer in the United States.
- Provide statistics on cancer incidence, survival, and mortality in the United States.
- 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 SEER Program collects and publishes data on cancer incidence and survival from 15 population-based cancer registries and 3 supplemental registries. SEER coverage includes 25.6% of African Americans, 24.9% of whites, 38.4% of Hispanics, 30.6% of American Indians and Alaska Natives, 50.4% of Asians, and 66.5% of Hawaiian/Pacific Islanders. 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 population coverage is noted below—

- The Detroit Registry covers 41% of Michigan's population.
- The Seattle-Puget Sound Registry covers 69% of Washington's population.
- The Greater Bay Area registries (San Francisco-Oakland Registry and San Jose-Monterey Registry) cover 19% of California's population.
- The Los Angeles County Registry covers 28% of California's population.
- The Alaska Native Tumor Registry covers 16% of Alaska's population.
- The Arizona Indians Registry covers 5% of Arizona's population.

Since 2001, NCI funding for Kentucky, Louisiana, New Jersey, and the remainder of California has provided resources for these registries to meet the requirements of the SEER Program regarding the metrics of completeness of case ascertainment, active patient follow-up, timeliness, and data quality. In 2012, Greater Georgia (the parts of Georgia not included in Atlanta and rural Georgia) was added to the SEER Program, with data retroactive to 2000. Information on more than 3 million *in situ* and invasive cancer cases is included in the SEER database, and approximately 170,000 new cases are added each year within SEER coverage areas.

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.
- Cancer statistics fact sheets.
- A restricted access data set for researchers through the National Center for Health Statistics' Research Data Center.

- A comprehensive overview of the most recent incidence, mortality, prevalence, lifetime risk, and survival statistics. The *Cancer Statistics Review* reports and summarizes the key measures of cancer's impact on the U.S. population.

Reference

1. Hankey BF, Ries LA, Edwards BK. The Surveillance, Epidemiology, and End Results Program: a national resource. *Cancer Epidemiology, Biomarkers and Prevention* 1999;8(12):1117–1121.

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 states.

Through its Vital Statistics Cooperative Program, NCHS cooperates with state vital statistics offices to develop and recommend standard forms for data collection and 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. 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. These data are obtained through the NCHS application process.

The NCHS issues a public use data set each year for additional analyses. The public use data set no longer contains geography as a variable. However, a data file containing information on geography is obtained for this report through a special request to the National Association of Public Health Statistics and Information Systems (the organization representing state vital registration systems) and to the NCHS.

Data Sources

Incidence Data

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.

Mortality Data

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 the National Vital Statistics System.

Population Denominator Data

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

Population estimates used in the calculation of Puerto Rico incidence and death rates are sex-specific, are from the 2010 U.S. Census, and are not available by race or ethnicity.

Incidence Data Sources

Data from the registries participating in the National Program of Cancer Registries (NPCR) were reported to the Centers for Disease Control and Prevention (CDC) as of November 30, 2017. Data from registries in the Surveillance, Epidemiology, and End Results (SEER) Program were reported to the National Cancer Institute (NCI) as of November 1, 2017, and made available through the SEER Program limited-use data file released in April 2018. Data from California, Georgia, Kentucky, Louisiana, and New Jersey (states that are supported by both NPCR and SEER) are presented as reported to CDC as of November 30, 2017.

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. 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 uniformity 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

NPCR and SEER cancer registries consider as reportable 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. 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.

Impact 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 as it was submitted to CDC. Incidence rates on this website may differ from those published by the SEER program for Louisiana, because the SEER program used only the first six months of incidence data for 2005 coupled with half of the population estimate for July 1, 2005, to calculate cancer incidence. Rates for the U.S. Census divisions and regions that include these states are calculated based on the data as submitted to CDC.

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/WHO 2008 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.

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 Sites" category.

Unknown Sex, Age, or Race

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

Nonmalignant Brain and CNS Tumors

Incidence data on nonmalignant primary brain and central nervous system (CNS) tumors are available on this website. 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]

References

2. Fritz A, Ries LAG. *The SEER Program Code Manual, Third Edition.* Bethesda (MD): National Cancer Institute; 1998.

3. Thornton ML, (ed). *Standards for Cancer Registries Vol II: Data Standards and Data Dictionary, Record Layout Version 14, 18th edition.* Springfield (IL): North American Association of Central Cancer Registries; 2013.

4. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, Whelan S. *International Classification of Diseases for Oncology, Third Edition.* Geneva, Switzerland: World Health Organization; 2000.

5. Hematopoietic codes based on *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (2008).* Accessed on April 24, 2016.

6. 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-oncology* 2005;7(2):196–201.

Mortality Data Sources

How Mortality Data Are Collected

Cancer mortality statistics are based on information from all death certificates filed in the 50 states and the District of Columbia 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]

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. Updates to the ICD in 2012 do not have a significant impact on the data on this website.

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 on this website were coded by procedures outlined in annual issues of the *NCHS Instruction Manuals.*

Underlying Cause of Death

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.[2] [3]

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 January 27, 2003. 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, there are several codes for mesothelioma in ICD-10 (depending on the primary site). However in ICD-O-3, one code captures all of the primary sites that mesothelioma affects.

Death Rates for Kaposi Sarcoma

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

Mortality Data Submission Process

Unlike incidence data, mortality data for a calendar year are not updated after the data file is released. All states and the District of Columbia submitted part or all of their 2015 mortality data in electronic data files to NCHS. Mortality data for the entire United States refer to deaths that occurred within the United States; data for geographic areas are by the decedent's place of residence.

References

1. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. Geneva, Switzerland: World Health Organization; 1992.

2. Chamblee RF, Evans MC. TRANSAX: the NCHS system for producing multiple cause-of-death statistics, 1968–78. Vital and Health Statistics, Series 1 1986;(20):1–83.

3. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. American Journal of Epidemiology 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, and sex-specific county population estimates aggregated to the state or metropolitan-area level. 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 under a collaborative arrangement between the U.S. Bureau of the Census (Census Bureau) and CDC's National Center for Health Statistics with support from NCI through an interagency agreement.

NCI's modifications to the population estimates are documented in Population Estimates Used in NCI's SEER*Stat Software. Several modifications pertain to the grouping of specific counties needed to assure the compatibility of all incidence, mortality, and population data sets. 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 used in the calculation of Puerto Rico incidence and death rates are sex-specific, are from the 2010 U.S. Census, and are not available by race or ethnicity.

Population Estimates

In general, July 1 population estimates are used to calculate annual incidence and death rates because these estimates are considered to reflect the average population of a defined geographic area for a calendar year. However, 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.

For these states, 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 the following eight states: Texas, Arkansas, Louisiana, Mississippi, Alabama, Tennessee, Georgia, or Florida. The evacuee population was included in the 2005 incidence rates since all of the relocation states met the USCS publication criteria.

U.S. Cancer Statistics Publication Criteria

Cancer incidence data that appear on this website are from central cancer registries that have high-quality cancer incidence data. The following are U.S. Cancer Statistics publication criteria—

- No more than 5% of cases are ascertained solely on the basis of a death certificate.
  - A measure of the completeness of case ascertainment is the proportion of cases ascertained solely on the basis of a death certificate, with no other information on the case available after the registry has completed a routine procedure known as "death clearance and followback."[1] [2] [3]

- No more than 3% of cases are missing information on sex.
- No more than 3% of cases are missing information on age.
- No more than 5% of cases are missing information on race.
- At least 97% of the registry's records passed a set of single-field and interfield computerized edits.
  - *Computerized edits* are computer programs that test the validity and logic of data components. For example, if (a) a patient received a diagnosis of cancer in 1999, (b) the patient's age was reported as 80 years, and (c) the patient's year of birth was reported as 1942, a computerized edit could, without human intervention, identify these components as incompatible. The computerized edits applied to the data in this report were designed by the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program for use by SEER registries. During the 1990s, these edits were expanded and incorporated into North American Association of Central Cancer Registries (NAACCR) standards and into the NPCR–EDITS software designed and maintained by CDC.

After years of analyzing completeness of case ascertainment, CDC has determined that NPCR registries consistently deliver high-quality, complete data.[4–7] Completeness of case ascertainment calculations have been discontinued as a measure of eligibility for publication. The data quality criteria—missing/unknown data, death-certificate-only percentage, duplicate rate, and percentage of records passing edits—will continue to be used in determining meeting or not meeting publication criteria. Even though the completeness estimate will no longer be a criterion for USCS, it will continue to be used to monitor and evaluate progress in meeting NPCR Program Standards.

Because some cancer patients receive diagnostic or treatment services at more than one reporting facility, cancer registries perform a procedure known as "unduplication" to ensure that each cancer case is counted only once.[1]

See central cancer registries that met USCS publication criteria.

References

1. Havener L. Standards for Cancer Registries, Vol III: Standards for Completeness, Quality, Analysis, and Management of Data. Springfield (IL): North American Association of Central Cancer Registries; 2004.

2. Menck HR, Phillips JL. Central cancer registries. In: Hutchinson CL, Menck HR, Burch M, Gottschalk R, editors. Cancer Registry Management: Principles and Practice. Second edition. Dubuque (IA): Kendall/Hunt Publishing Company; 2004.

3. Seiffert JE, Hoyler SS, McKeen K, Potts M. Casefinding, abstracting, and death clearance. In: Menck HR, Smart C, editors. Central Cancer Registries: Design, Management, and Use. Chur, Switzerland: Harwood Academic Publishers; 1994.

4. Tucker TC, Howe HL, Weir HK. Certification of population-based cancer registries. Journal of Registry Management 1999;26(1):24–27.

5. Tucker TC, Howe HL. Measuring the quality of population-based cancer registries: the NAACCR perspective. Journal of Registry Management 2001;28(1):41–44.

6. Johnson CJ, Weir HK, Mariotto AB, Nishri D, Wilson R, Copeland G, Lake A, Firth R, Wohler B, Wu XC, Schymura M, Hofferkamp J, Sherman R, Kohler B (eds). Cancer in North America: 2009-2013 Volume Four: Cancer Survival in the United States and Canada 2006-2012. Springfield, IL: North American Association of Central Cancer Registries, Inc. June 2016.

7. Howe HL. Conclusions of the Workgroup for High-Quality Criteria for Data Use: The NAACCR Narrative. Springfield (IL): North American Association of Central Cancer Registries; 2001.

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 November 2017. Criteria must be met for all diagnosis years for the combined 2011 to 2015 data.

- **2011 to 2015:** 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, after manual review of Nevada's data. 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 met the publication criteria except Mississippi, after manual review of the District of Columbia's data. Counts and rates cover approximately 99% of the U.S. population.
- **2001:** All registries met the publication criteria except data are not available for Mississippi. Counts and rates cover approximately 99% of the U.S. population.
- **2000:** All registries met the publication criteria except Arkansas; data are not available for Mississippi and South Dakota. Counts and rates cover approximately 98% of the U.S. population.
- **1999:** All registries met the publication criteria except Arkansas; data are not available for Mississippi and South Dakota. Counts and rates cover approximately 98% of the U.S. population.

U.S. Census Regions and Divisions That Met U.S. Cancer Statistics Publication Criteria

Publication criteria were assessed on data submitted to CDC's National Program of Cancer Registries and the National Cancer Institute's Surveillance, Epidemiology, and End Results Program in November 2017. Criteria must be met for all diagnosis years for the combined 2011–2015 data.

For information about census divisions and regions, see Geographic Terms and Concepts – Census Divisions and Census Regions.

- **2011 to 2015:** All U.S. Census regions and divisions met the reporting criteria.
- **2015:** All U.S. Census regions and divisions met the reporting criteria.
- **2014:** All U.S. Census regions and divisions met the reporting criteria.
- **2013:** All U.S. Census regions and divisions met the reporting criteria.
- **2012:** All U.S. Census regions and divisions met the reporting criteria.
- **2011:** All U.S. Census regions and divisions met the reporting criteria.
- **2010:** All U.S. Census regions and divisions met the reporting criteria.
- **2009:** All U.S. Census regions and divisions met the reporting criteria.
- **2008:** All U.S. Census regions and divisions met the reporting criteria.
- **2007:** All U.S. Census regions and divisions met the reporting criteria.
- **2006:** All U.S. Census regions and divisions met the reporting criteria.
- **2005:** All U.S. Census regions and divisions met the reporting criteria.
- **2004:** All U.S. Census regions and divisions met the reporting criteria.
- **2003:** All U.S. Census regions and divisions met the reporting criteria.
- **2002:** The East South Central division and the South region did not meet the reporting criteria. All other U.S. Census regions and divisions met the reporting criteria.
- **2001:** The East South Central division and the South region did not meet the reporting criteria. All other U.S. Census regions and divisions met the reporting criteria.
- **2000:** The East South Central division and the South region did not meet the reporting criteria. All other U.S. Census regions and divisions met the reporting criteria.
- **1999:** The East South Central and West South Central divisions and the South region did not meet the reporting criteria. All other U.S. Census regions and divisions met the reporting criteria.

Criteria for Reporting Age-Adjusted Cancer Incidence Rates for U.S. Census Regions and Divisions

The annual age-adjusted cancer incidence rates for some U.S. Census regions and divisions are not available because the data from the cancer registries of some states in those regions or divisions do not meet the eligibility criteria for inclusion in this report. In contrast, the annual age-adjusted cancer death rates are available for all states in every Census region or division. However, the age-adjusted incidence rate for Census regions or divisions in which the data of less than 100% of the cancer registries meet eligibility criteria can be estimated by assuming that the incidence-to-mortality ratio for states *without* eligible cancer registry data in that Census region or division equals the incidence-to-mortality ratio for states *with* eligible cancer registry data in that Census region or division.

Let

M_1 = age-adjusted death rate in states with eligible cancer registries

M_0 = age-adjusted death rate in states without eligible cancer registries

I_1 = age-adjusted incidence rate in states with eligible cancer registries

I_0 = age-adjusted incidence rate in states without eligible cancer registries (incidence data are not available)

P_1 = Proportion of the population in the Census region or division that resides in states with eligible cancer registries

$R = (M_0/M_1)$

$\hat{I}_{total}$ = age-adjusted incidence rate for the entire Census region or division

where “eligible” refers to the state and metropolitan area cancer registries that meet this report’s data quality criteria for all invasive cancer sites combined.

Since we are assuming that $\frac{I_1}{M_1} = \frac{I_0}{M_0}$, the estimate of the age-adjusted incidence rate for states without eligible cancer registries is $I_0 = I_1(\frac{M_0}{M_1}) = I_1 R$. Thus, an estimate of the age-adjusted incidence rate for 100% of the Census region or division is computed as the following weighted average—

$$\hat{I}_{total} = P_1 I_1 + (1 - P_1) I_0 = P_1 I_1 + (1 - P_1) I_1 R = I_1[P_1 + (1 - P_1) R]$$

As an example, consider invasive female breast cancer in a hypothetical Census region with seven states. Incidence data for five states that cover 86.3% of the population (P_1 = 0.863) are eligible for inclusion in the calculation of the regional incidence rate; data for two states are not eligible. The female breast cancer death rate for the five eligible states is $M_1 = \frac{27.3}{10^5}$, and the rate for the two ineligible states is $M_0 = \frac{27.7}{10^5}$.

The age-adjusted incidence rate for states with eligible cancer registries is $I_1 = \frac{145.1}{10^5}$.

The age-adjusted incidence rate for female invasive breast cancer in the entire Census region (in other words, corrected for the data not available from the ineligible registries) is

$$\hat{I}_{total} = \frac{145.1}{10^5} * [0.863 + 0.137(\frac{27.7}{27.3})] = \frac{145.39}{10^5}$$

The underlying assumptions for this method are that the age-adjusted death rates for states with and without eligible cancer registries are accurate and that the incidence-to-mortality ratio for states without eligible cancer registries in that Census region or division equals the incidence-to-mortality ratio for states with eligible cancer registries in that Census region or division.

For each Census region or division in which less than 100% of the registries provided data eligible for this report, we used the above-described method to estimate the age-adjusted incidence rates ($\hat{I}_{total}$) for the six major cancer sex-site groups: breast (female only), prostate, male and female colorectal, and male and female lung and bronchus. If the estimate of the age-adjusted incidence rate for each of the six cancer sites for that Census region or division falls within the confidence interval of the observed age-adjusted incidence rate for states with eligible cancer registries, then the observed age-adjusted incidence rates for *all* cancer sites are published. If one or more of the six estimates of age-adjusted incidence rates falls outside the confidence interval, then the observed age-adjusted cancer incidence rates are not reported for that U.S. Census region or division.

We emphasize, however, that all cancer incidence rates in this report are based exclusively on data obtained from states with eligible cancer registries and are not the estimates of the age-adjusted incidence rates calculated using the methods described above.

Statistical Methods

Incidence and Death Rates

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

Confidence Intervals

Confidence intervals reflect the range of variation in the estimation of the cancer rates. The width of a confidence interval depends on the amount of variability in the data.

Relative Cancer Survival

Surveillance of cancer incidence and survival are essential in monitoring and understanding CDC's efforts to support the needs of cancer survivors.

Cancer Prevalence

Prevalence helps identify the level of burden of disease on the population and health care system. It is a function of both incidence and survival.

Suppression of Rates and Counts

When the numbers of cases or deaths used to compute rates are small, those rates tend to have poor reliability. 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 or eliminating the risk of disclosing their identity.

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 vs. Age-Adjusted Rates

Crude rates are influenced by the underlying age distribution of the state's population. Even 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 categories) 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 population used to age-adjust the rates in this report is the 2000 U.S. standard population.[1] [2] On this website, the 2000 U.S. standard population is based on the proportion of the 2000 population in 19 specific age groups (younger than 1 year, 1–4 years, 5–9 years, 10–14 years, 15–19 years, … 85 years and older); except for Puerto Rico, where it is based on 18 specific age groups (0–4 years, 5–9 years, 10–14 years, 15–19 years, … 85 years and older); the proportions of the 2000 population in these age groups serve as weights for calculating age-adjusted incidence and death rates. Cancer death rates on this website may differ slightly from those published by the National Center for Health Statistics (NCHS) because NCHS uses age groups as recommended by the U.S. Department of Health and Human Services in its adjustment of death rates. In addition, 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, 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.[2]

References

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

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

Confidence Intervals

Confidence intervals reflect the range of variation in the estimation of the cancer rates. The width of a confidence interval depends on the amount of variability in the data. Sources of variability include the underlying occurrence of cancer as well as uncertainty about when cancer is detected and diagnosed, when a death from cancer occurs, and when the data about the cancer are sent to the registry or the state health department.

In any given year, when large numbers of a particular cancer are diagnosed or when large numbers of cancer patients die, the effects of random variability are small compared with the large numbers, and the confidence interval will be narrow. With rare cancers, however, the rates are small and the chance occurrence of more or fewer cases or deaths in a given year can markedly affect those rates. 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 is capable of estimating separate parameters that represent contributions to the rate from various population risk factors, the effects of cancer control interventions, and other attributes of the population risk profile in any particular year.

Modified Gamma Intervals

Confidence intervals that are expected to include the true underlying rate 95% of the time are used on this website and 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

The use of overlapping confidence intervals to test for statistically significant differences between two rates presented on this website is discouraged because 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 on this website, numerators and denominators are large and standard errors are therefore small, resulting in statistically significant differences that may be so small as to lack importance 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. *Statistical Methods in Medical Research* 2006;15(6):547–569.
4. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Statistics in Medicine* 1997;16(7):791–801.
5. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician* 2001;55(3):182–186.

Relative Cancer Survival

Surveillance of cancer incidence and survival are essential in monitoring and understanding CDC's efforts to support the needs of cancer survivors, estimated to be 15.1 million in 2015.[1]

Definition and Calculation of Relative Cancer Survival

The *relative cancer survival rate* 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. The relative cancer survival rate 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.[1] 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.

Cancer incidence data submitted to National Program of Cancer Registries (NPCR) as of November 30, 2017, were used to create a data set in SEER*Stat for this analysis.[2] The data set included data from 39 NPCR central cancer registries that met the United States Cancer Statistics (USCS) publication criteria for all years 2001 through 2014 and that conducted linkage with the National Death Index and/or active patient follow-up for all years 2001 through 2014. These registries include Alabama, Alaska, Arizona, Arkansas, California, Colorado, Delaware, Georgia, Idaho, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Minnesota, Mississippi, Missouri, Montana, Nebraska, New Hampshire, New Jersey, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, Wisconsin, and Wyoming. These data cover 81% 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 2014. Cases diagnosed in 2015 do not have adequate follow-up time to be included in the analysis.
- The age of the case was known and was 0 through 99 years.
- The sex of the case was known.
- The case was not identified solely on the basis of a death certificate or autopsy.

Analytic Methods

Survival time in months for each case was calculated. Date of start of follow-up (month, day, and year) was set to date of diagnosis. Date of last follow-up (month, day, and year) was set to date of death if the case was matched to the state death files, to the National Death Index, or to date of last contact (if case was actively followed). Cases not linking to the state death files or to the National Death Index were presumed to be alive, and the date of last follow-up was set to December 31, 2014. Where day or month for date of diagnosis, date of death, or date of last contact were missing, the full date was imputed using a standard algorithm.[3] Cases that survived past the maximum age (99 years) were censored at age 99. Observed all-cause survival by sex and race (white, black, and all races combined) for individuals with any cancer and for individuals with 23 common cancer sites was then calculated using the actuarial life table method.[4] Cases with multiple primary cancers were included in the dataset, although only the first primary cancer was included in calculating relative survival for all cancer sites combined. Where a patient had multiple primary cancers of **different** sites, each cancer was included in calculating cancer-specific relative survival. Where 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, but if diagnosed at **different** ages, each cancer was included in the calculation.[5]

Expected all-cause survival for the general population by sex and race (white, black and all races combined) were obtained using annual U.S. life tables provided by the National Center for Health Statistics and modified by SEER. The life tables were embedded in SEER*Stat. See Expected Survival Life Tables for more information.

Relative cancer survival was then calculated using the Ederer II method[6] for all cancer sites combined and for 23 common cancer sites by sex, race (all races combined, white, black, and other), and age group (younger than 45, 45 to 54, 55 to 64, 6 to 74, 75 or older). The other races group contains Indian Health Service-linked American Indian, Alaska Native, and Asian/Pacific Islander cases. See Measures of Cancer Survival for more information.

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

References

1. Noone AM, Howlader N, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975–2015, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER Web site, April 2018.

2. NPCR survival analytical database, Submission 2016 November, 2001–2013, 34 NPCR states.

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

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

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

6. 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. Available from: http://surveillance.cancer.gov/reports/.

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. *Journal of Registry Management* 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 14 years).

Cancer incidence data submitted to National Program of Cancer Registries (NPCR) as of November 30, 2017, were used to create a data set in SEER*Stat for this analysis.[2] The data set included data from 39 NPCR central cancer registries that met the United States Cancer Statistics (USCS) publication criteria for all years 2001 through 2014 and that conducted linkage with the National Death Index and/or active patient follow-up for all years 2001 through 2014. These registries include Alabama, Alaska, Arizona, Arkansas, California, Colorado, Delaware, Georgia, Idaho, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Minnesota, Mississippi, Missouri, Montana, Nebraska, New Hampshire, New Jersey, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, Wisconsin, and Wyoming. These data cover 81% 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 2014. Cases diagnosed in 2015 do not have adequate follow-up time to be included in the analysis.
- The age of the case was known and was 0 through 99 years.
- The sex of the case was known.
- The case was not identified solely on the basis of a death certificate or autopsy.

Because NPCR data are available from 2001, 14-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 14 years).

In this report, the limited-duration prevalence was calculated using SEER*Stat software. It estimates, among the people diagnosed with cancer in the last 5 or 14 years, the proportion who were still alive on January 1, 2015.[1,2] The date of start of follow-up (month, day, and year) was set to the date of diagnosis. The date of last follow-up (month, day, and year) was set either to the date of last contact (if the case was actively followed) or to the date of death if the case was matched to the state death files or to the National Death Index. Cases not linking to the state death files or to the National Death Index were presumed to be alive on the prevalence date.

For patients diagnosed with multiple tumors, prevalence calculations include the first tumor of each cancer type in the previous *x* years (where *x* = 5 or 14 in this report). For example, a woman was diagnosed first with thyroid cancer 9 years ago and then breast cancer 3 years ago. The thyroid cancer would contribute to the 14-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 for breast cancer, but not to the 14-year limited-duration prevalence estimate for breast cancer because the woman is already counted in this estimate for thyroid cancer.

NPCR prevalence proportions were calculated for each combination of age, sex, and race group. For this report, race was categorized as white, black, and other races. The other races group contains Indian Health Service-linked American Indian, Alaska Native, and Asian/Pacific Islander cases. Cases with unknown race were combined with white race. Then, cancer prevalence counts at January 1, 2015, for the U.S. population were estimated by multiplying the age-, sex-, and race-specific NPCR prevalence proportions by the corresponding U.S. population estimates based on the average of the 2014 and 2015 population estimates from the U.S. Census Bureau.[3] U.S. cancer prevalence counts for all races combined were estimated by summing the counts for whites/unknown, blacks, and other races.

References

1. Surveillance Research Program, National Cancer Institute SEER*Stat software version 8.3.45.
2. 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.
3. SEER*Stat Database: NPCR Prevalence Database, 2016 November (34 states, counties in Kansas and Minnesota excluded, cutoff January 1, 2014)

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 if 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 as large as the rate itself. Similarly, a case count below 16 results in the width of the rate's 95% confidence interval 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 or eliminating the risk of disclosing their identity.[2] The cell suppression threshold value of 16, which was selected to reduce misuse and misinterpretation of unstable rates and counts in this report, is more than sufficient to protect patient confidentiality.[PDF-324KB]

Suppression for Bar Graphs

Because the incidence and death rates shown in the state-, sex-, and race-specific bar graphs are presented in rank order, we applied a criterion for suppressing data in addition to the threshold value of 16 cases. In these figures, incidence and death rates are not ranked or shown for any population groups of fewer than 50,000 people.

Suppression for Other Reasons

While data meet the USCS publication criteria, a central cancer registry may suppress its data for various reasons. For example, a state may have racial and ethnic groups (American Indian/Alaska Native, Asian/Pacific Islander, Hispanic) where the algorithms to correct for unknown race or ethnicity may not function properly. In these circumstances, data are suppressed upon the state's request.

U.S. Census Regions and Divisions

Rates for U.S. Census regions and divisions were calculated by aggregating data reported from the states in each region and division. Only data from state registries that met USCS publication criteria were included in calculations of incidence rates for U.S. Census regions and divisions. Thus, where data for some states are excluded there is a potential for bias in the incidence rates for Census regions and divisions. We estimated cancer rates for Census regions or divisions with ineligible cancer registries by assuming that the incidence-to-mortality ratio in the portion of the region or division that was covered by eligible registries was the same as the incidence-to-mortality ratio in the portion that was not covered by eligible cancer registries. The age-adjusted incidence rates for U.S. Census regions and divisions are presented only if—

1. At least 80% of the population for the Census region or division was covered by cancer registries that met USCS publication criteria.

2. The 95% confidence intervals around the observed age-adjusted regional or division incidence rates based on data from eligible registries for each of six major cancer sites (prostate, female breast, male colorectal, female colorectal, male lung and bronchus, female lung and bronchus) included the estimate of the regional or division rate calculated using the specified criteria.

This website presents the observed age-adjusted incidence rates for all U.S. Census regions and divisions. Case counts for U.S. Census regions and divisions are presented if all state cancer registries in the region or division met the criteria for inclusion, unless the count for one state in the region or division is suppressed due to a count below 16.

U.S. State and County Data

Cancer incidence rates are presented for each county or county equivalent as available. County data are not available from Kansas and Minnesota because of state legislation and regulations which prohibit the release of county-level data to outside entities. Data are suppressed in accordance with the rules outlined above.

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 USCS publication criteria and 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.

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 the Data

Incidence Data

Each year when United States 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.

Mortality Data

Cancer mortality statistics on this website are influenced by the accuracy of information on the death certificate.

Race and Ethnicity in Cancer Data

In cancer incidence, race and ethnicity information is abstracted from medical records and grouped into categories. 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, in the absence of an informant, on the basis of observation.

Guidance for Comparing States' Cancer Data

Careful consideration is needed in interpreting and comparing rankings of state cancer rates. A natural reaction of some readers when looking at figures that rank their state's cancer rates is to seek explanations as to why their state has higher incidence or death rates for some cancers than other states or than the national average. For example, some may be alarmed that exposure to environmental carcinogens may be responsible when in fact there are several other more likely explanations.

Interpreting the Incidence Data

Change to the 2000 U.S. Standard Population

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] The change to the 2000 U.S. standard population updated the calculation of age-adjusted rates to reflect more closely the current age distribution of the U.S. population. Because of the aging of the U.S. population, the 2000 U.S. standard population gives more weight to older age categories than the 1940 and 1970 standard populations did.[2]

Because the incidence of cancer increases with age, the change to the 2000 U.S. standard population resulted in higher incidence rates for most cancers. The data on this website 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 on this website may differ slightly from those published by individual states.

Statistical Bias

Statistical bias can arise if, within a region, division, or country, the sub-area for which data are available has rates that differ substantially from the rates in the sub-area for which data are not available. Because of bias, rates for a U.S. Census region, U.S. Census division, or the country may not meet statistical criteria for inclusion. It is possible to have some statistical bias even if the percentage of coverage is high and large numbers of cases are recorded. Where coverage is less than 100%, merely increasing the percentage of the population covered may not reduce statistical bias unless the covered population is similar to the uncovered population in terms of cancer rates or proportions.

Registries' Data Quality

Data quality is evaluated routinely by CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program.[3] [4] Some evaluation activities are conducted intermittently 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. 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 primary cancer site to another.

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 differed by 30 days: NPCR allowed an interval of 23 months after the close of the diagnosis year (data submission by November 30, 2017), and SEER allowed an interval of 22 months after the close of the diagnosis year (data submission by November 1, 2017).

Reporting Delays

Delays in reporting cancer cases can affect the timely and accurate calculation of cancer incidence rates.[6] 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 revise and 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 on this website. 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 cancer, melanoma of the skin). Methods to adjust incidence rates for reporting delay were not applied to the data in this report.[6]

Continual Data Updates

Each year, state cancer registries submit data for a new diagnosis year to CDC or NCI, plus an updated version of previous years' data. Federal agencies in turn update their cancer incidence statistics with each data submission and document the states' date of data submission whenever the data are published. These continual updates by state and

federal agencies illustrate the dynamic nature of cancer surveillance and the attention to detail that is characteristic of cancer registries. Each year when United States 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.

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.[7] [8] [9] [10] Differences may arise because of differences in sociodemographic characteristics of the population (age, race and ethnicity, geographic region, urban or rural residence), screening use, health-related behaviors (for example, tobacco use, diet, physical activity), exposure to cancer-causing agents, or factors associated with the registries' operations (completeness, timeliness, specificity in coding cancer sites). Cancer researchers are investigating variability associated with known factors that affect cancer rates and risks by using 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 and Health Statistics, Series 4.* 1998;(30):I–VI, 1–37.

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

3. Fritz A. The SEER Program's commitment to data quality. *Journal of Registry Management* 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. *Journal of Registry Management* 2001;28(3):113–120.

5. Thoburn KK, German RR, Lewis M, Nichols PJ, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer* 2007;109(8):1607–16.

6. Clegg LX, Feuer EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. *Journal of the National Cancer Institute* 2002;94(20):1537–1545.

7. Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System Operational and User's Guide. Version 3.0.*[PDF-986KB] Atlanta (GA): Centers for Disease Control and Prevention; 2005.

8. 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.

9. Howe HL, Keller JE, Lehnherr M. Relation between population density and cancer incidence, Illinois, 1986–1990. *American Journal of Epidemiology* 1993;138(1):29–36.

10. Wingo PA, Jamison PM, Hiatt RA, Weir HK, Gargiullo PM, Hutton M, Lee NC, Hall HI. 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 and Control* 2003;14(2):175–193.

Interpreting Mortality Data

Cancer mortality statistics on this website are influenced by the accuracy of information on the death certificate. 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%–93%, and their positive predictive value was 85%–96%.[1] [2] [3] However, these studies are limited by selection bias, and less 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 <50%, some were 95% or greater, and notable differences between confirmation and detection rates were observed. For some of the most commonly diagnosed cancers in the U.S. (for example, prostate, breast, and lung and bronchus), confirmation or detection rates were 95% or greater. 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 re-engineered to collect data electronically. These systems will accommodate better certificate revisions, special studies or projects, and linkage with other health promotion programs. With regard to mortality data, handbooks have been revised for professionals who complete death certificates.

References

1. Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. *New England Journal of Medicine* 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. *American Journal of Epidemiology* 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. *Bulletin of the New York Academy of Medicine* 1982;58(9):778–794.

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

5. German RR, Fink AK, Heron M, Stewart SL, Johnson CJ, Finch JL, Yin D; the Accuracy of Cancer Mortality Study Group. The accuracy of cancer mortality statistics based on death certificates in the United States. *Cancer Epidemiology* 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.[1] 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 (for example, Hispanic origin), 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.

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, in the absence of an informant, 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] The net effect of misclassification is greatest for American Indians/Alaska Natives; misclassification is smaller for Asians/Pacific Islanders and Hispanics and minimal for blacks and whites. Therefore, incidence and/or mortality data published in this report may be underestimated for Asians/Pacific Islanders, American Indians/Alaska Natives, and Hispanics, possibly due to racial and Hispanic origin misclassification. CDC's National Center for Health Statistics is working with states to improve the reporting of race and ethnicity on death certificates.

This website presents cancer incidence and mortality data for all races combined and by race and ethnicity (Hispanics). Data for Asians/Pacific Islanders and American Indians/Alaska Natives are presented only for the nation and for states with at least 50,000 population because of concerns regarding possible misclassification of race data and the relatively small sizes of these populations in the United States.

Asians/Pacific Islanders

Although state cancer registries have designated codes for race that allow them to document the occurrence of cancer in 23 Asian/Pacific Islander subpopulations,[2] the subpopulations are grouped into a single Asian/Pacific Islander category because of small numbers and concerns regarding possible misclassification of race data.

Studies show excellent agreement (k=0.90) between Asian/Pacific Islander race in Surveillance, Epidemiology, and End Results (SEER) registry data and self-reported data from the U.S. Census.[5] Studies are underway to examine the misclassification of race for Asian/Pacific Islander subpopulations in cancer registries.[6] [7] [8] 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/Pacific Islander Identification Algorithm (NAPIIA) version 1.2.

The following NPCR registries opted not to present state- and county-specific Asian/Pacific Islander counts and rates: Delaware, Kansas, and Kentucky. The national rates presented include data for these registries.

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

Hispanics

The overall agreement between Hispanic ethnicity collected by SEER registries and self-reported ethnicity from the U.S. Census was substantial (k=0.61). Hispanics were found to be underclassified in the SEER data compared to self-reports.[5] Nearly all NPCR and SEER registries assigned Hispanic ethnicity through the standardized use of the NAACCR Hispanic Identification Algorithm (NHIA) version 2 (NHIAv2.2.1). After applying the NHIAv2, cases not classified as Hispanic are classified as non-Hispanic, leaving no cases with unknown Hispanic status.

The following NPCR registries opted not to present state- and county-specific, NHIA-classified Hispanic counts and rates for all years: Delaware, Kentucky, and Massachusetts. The national rates presented include data for these registries.

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

American Indians/Alaska Natives

Studies that estimate misclassification among American Indians/Alaska Natives using cancer registry data report these rates are underreported by 40%–57%, depending on the region of the country.[9] [10] [11] These studies have linked cases with Indian Health Service (IHS) administrative records; IHS provides medical services to American Indians/Alaska Natives who are members of federally recognized tribes. IHS coverage of these populations varies by region, does not include American Indians/Alaska Natives who are members of non-federally recognized tribes, and underrepresents those who live in certain urban areas. American Indians/Alaska Natives who live outside of service counties may continue to receive IHS services or may have received services before moving. To address American Indian/Alaska Native misclassification in cancer registry data, selected NPCR and all SEER registries linked their data to the IHS administrative records database for cases diagnosed from 1995 to 2014 and 1988 to 2014, respectively. Results of the linkage were captured in a new data element, IHS Link (NAACCR data element 192),[2] that was sent back to state cancer registries. In turn, the state cancer registries submit IHS Link to CDC or NCI and the federal agencies use this variable in conjunction with the race as it is coded in the medical records to assign a recoded race variable. For Alaska, IHS Link was not used to determine race.

Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York have opted not to present state- and county-specific American Indian/Alaska Native counts and rates. The national rates presented include data for these registries.

Studies have found that racial misclassification contributes to lower death rates and lower cancer incidence rates among the American Indian/Alaska Native population. Based on a comparison of race reported on death certificates and self-reported race, record-by-record agreement was only 55% for American Indians.[4] When adjusted for this misclassification, the age-adjusted American Indian/Alaska Native death rate was 11% greater than the age-adjusted rate of the white population; prior to the adjustment, the age-adjusted death rate was 15% lower than the rate for the white population.[4] National death counts and rates for American Indians/Alaska Natives are based on data obtained from all 50 states and the District of Columbia. Classification as American Indian/Alaska Native is obtained from information on the death certificate.

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.[PDF-3.4MB] *National Vital Statistics Reports* 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.[PDF-3.1MB] *Vital and Health Statistics* 2016;2(172).

5. Clegg LX, Reichman ME, Hankey BF, Miller BA, Lin YD, Johnson NJ, Schwartz SM, Bernstein L, Chen VW, Goodman MT, Gomez SL, Graff JJ, Lynch CF, Lin CC, Edwards BK. Quality of race, Hispanic ethnicity, and immigrant status in population-based cancer registry data: implications for health disparity studies. *Cancer Causes and Control* 2007;18(2):177–187.

6. NAACCR Race and Ethnicity Work Group. *NAACCR Asian Pacific Islander Identification Algorithm [NAPIIA v1.2.1]*[PDF-132KB]. Springfield (IL): North American Association of Central Cancer Registries; 2008.

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

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

9. 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.

10. Frost F, Taylor V, Fries E. Racial misclassification of Native Americans in a Surveillance, Epidemiology, and End Results cancer registry. *Journal of the National Cancer Institute* 1992;84(12):957–962.

11. 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. *Journal of Registry Management* 1998;25(1):17–20.

Guidance for Comparing States' Cancer Data

Note: For additional information on data interpretation, please refer to the USCS Technical Notes—Interpreting the Data.

Careful consideration is needed in interpreting and comparing rankings of state cancer rates.

A natural reaction of some readers when looking at figures that rank their state's cancer rates is to seek explanations as to why their state has higher incidence or death rates for some cancers than other states or than the national average. For example, some may be alarmed that exposure to environmental carcinogens may be responsible when in fact there are several other more likely explanations. The following points should be kept in mind when interpreting or comparing these rankings—

Differences Among Racial and Ethnic Populations

Some cancers have different cancer rates for different racial and ethnic populations. For example, breast cancer incidence rates are usually higher in white women than in women of other racial and ethnic populations, and prostate cancer incidence rates are higher in black men. Therefore, when comparing cancer rates across states, the racial makeup of the state's population should be taken into account and may be determined through the statistical adjustment of rates by race and ethnicity. However, presentation of rates for specific racial and ethnic populations may be preferable and is more easily understood by a lay audience.

Variations in Populations and Health Behaviors

Some differences in cancer rates among states may be explained by differences in known risk factors among the populations of those states. For example, one finds higher rates of lung cancer and other tobacco-associated cancers in states with higher rates of smoking. Although environmental carcinogens are responsible for some cancer cases, a majority of cases appear to be related to lifestyle factors such as smoking, and geographic variations in cancer rates are thought largely to reflect variations in these lifestyle factors.

Variations in Medical Care

Variations among states in medical care factors may also result in differences in cancer rates. In states 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 at times find tumors that grow so slowly that they would not otherwise be recognized in a person's lifetime. Therefore, the cancer incidence rate without additional information only tells part of the story.

Influence of Aging on Cancer Rates

The likelihood of being diagnosed with cancer increases steadily with age. These rates have been adjusted for age so that states 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 states.

Measuring Burden

The importance of cancer as a public health problem in a state is more a function of the absolute rate of cancer rather than the state's relative ranking in incidence or mortality. For example, Utah has proportionately fewer smokers 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 ranking in incidence compared with other states. Also, the true burden of cancer on the health care system and economy of a state is determined by the number of people diagnosed with or the number of people dying of cancer and not by the age-adjusted cancer rate. Therefore, the observation that the cancer rate in one state appears high compared with other states may obscure the fact that the absolute number of cases is not large.

Completeness of Cancer Incidence Data

States contribute cancer incidence data to these ranking figures if their registries collected 90% or more of the cancers. Because states vary in their completeness above 90%, rankings may vary to a minor extent because of differences in reporting completeness.

Random Factors and Cancer Rates

Even if registries were able to collect 100% of diagnosed cancer cases, there would still be some uncertainty in computed cancer rates because many factors contribute to the incidence and death rate in any given year or state, and some factors exhibit random behavior. Chance plays a role in determining if and when cancer develops in an individual, whether that cancer is detected, whether the information is entered into the cancer registry, and whether that cancer progresses and leads to death. For these reasons, the reported rates are expected to vary from year to year within a state even in the absence of a general trend. Caution is warranted, therefore, when examining cancer rates for a single year, and 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 state age-adjusted rates are available to help with interpreting the results. Because of the variation in the population sizes and number of reported cases and deaths across states, there is more uncertainty in the incidence and death rates for some states compared with others. The confidence intervals provide a measure of the variability in the rates and some perspective for making state-specific comparisons. It should be noted, however, that using overlapping confidence intervals to conclude that rates are not significantly different is not recommended. This is a conservative test because it fails to detect significant differences more often than does standard statistical hypothesis testing.

Public Health Importance

Another consideration when comparing differences between rates is their public health importance. For some rates in this report, numerators and denominators are large and the standard errors are small with the result that some statistically significant differences may be so small as to lack importance for decisions related to population-based public health programs.