U.S. Cancer Statistics Public Use Database Technical Documentation

U.S. Data
November 2023 Submission
Diagnosis Years 2001–2022

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
CENTERS FOR DISEASE
CONTROL AND PREVENTION

Contents

U.S. Cancer Statistics Public Use Database

Researchers can analyze high-quality cancer incidence data on the United States population. De-identified cancer incidence data are available to researchers free of charge in a public use database.

Cancer incidence data reported to CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) Program are combined to become U.S. Cancer Statistics (USCS), the official federal cancer statistics. The U.S. Cancer Statistics public use database includes cancer incidence and population data for all 50 states, the District of Columbia, and Puerto Rico, providing information on more than 39 million cancer cases since 2001. The database can be analyzed using software developed by NCI's SEER Program.

About the Database

The U.S. Cancer Statistics public use database includes cancer incidence and population data for all 50 states and the District of Columbia since 2001, providing information on more than 39 million cancer cases.

The database includes data by demographic characteristics (for example, age, sex, and race) and tumor characteristics (for example, year of diagnosis, primary tumor site, histology, behavior, and stage at diagnosis).

Hospitals, physicians, and laboratories across the nation report these data to central cancer registries supported by CDC and the National Cancer Institute (NCI). The databases are intended for researchers to conduct focused analyses beyond what is available through the U.S. Cancer Statistics Data Visualizations tool.

Researchers, public health professionals, clinicians, decision makers, and others can use these data to inform scientific inquiries, programs, and policies by identifying disparities in cancer burden, investigating trends and geographic distributions in cancer incidence, and evaluating and monitoring cancer prevention activities.

The current data come from the 2024 National Program of Cancer Registries (NPCR) and Surveillance, Epidemiology, and End Results (SEER) program submissions, which include cancer cases diagnosed from January 1, 2001 through December 31, 2022. Each year, NPCR- and SEER-supported central cancer registries submit data from a referent year to the close of the most current diagnosis year. The submitted data include information from previous years and are updated with information from the newly submitted records to ensure case completeness and high quality.

CDC and NCI support the data collection and quality standards in the North American Association of Central Cancer Registries (NAACCR) consensus documents. During data collection, CDC and NCI also apply additional rigorous quality control edits, data completeness evaluations, and data quality assessments. For a registry's data to be included in the U.S. Cancer Statistics public research data file, they must meet the U.S. Cancer Statistics publication standard.

Number of records in the database

The list below shows the number of cases available for the most recent U.S. Cancer Statistics data release.*

- All cases: 39,650,615 (includes benign and borderline brain and other nervous system tumors from 2004 onward)
- Malignant cases†: 35,707,286
- Malignant and *in situ* cases†: 38,512,343

*The following criteria apply to the U.S. Cancer Statistics public use database:

NPCR- and SEER-supported cancer registries report all incident cases coded as *in situ* (non-malignant) and invasive (malignant; primary site only), and non-malignant (including borderline and benign) central nervous system tumors according to the *International Classification of Diseases for Oncology, Third Edition* (ICD-O-3), with the following exceptions:

- *In situ* cancers of the cervix are not reported.
- Basal and squamous cell carcinomas of the skin are not reported, except when these occur on the skin of the genital organs.
- Additionally, *in situ* urinary bladder cancers were re-coded as invasive behavior.

†Malignant and *in situ* cases are defined using *Behavior code ICD-O-3*.

Suggested citations

Please use these standard citations for tables and figures when presented in presentations or publications.

For population coverage

Data are from population-based registries that participate in CDC's National Program of Cancer Registries and/or NCI's Surveillance, Epidemiology, and End Results Program and meet high-quality data criteria. Data from these registries from 2001 to 2022 cover approximately 99.9% of the U.S. population.

For age-adjusted rates

Rates are per 100,000 persons and are age-adjusted to the 2000 U.S. standard population (19 age groups – Census P25–1130).

For the database

National Program of Cancer Registries and Surveillance, Epidemiology, and End Results Program SEER*Stat Database: NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2022 Public Use Research Database, 2024 submission (2001–2022), U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June 2025. Available at www.cdc.gov/united-states-cancer-statistics/public-use/.

Central cancer registries by supporting program

NPCR funds 50 cancer registries: 46 states, the District of Columbia, Puerto Rico, the Pacific Island Jurisdictions, and the U.S. Virgin Islands. The SEER program collects and publishes data on cancer incidence and survival from population-based cancer registries in 21 U.S. geographic areas.

Central cancer registries submitting data to CDC and NCI during 2024 data collection

Alabama ..... CDC NPCR	Louisiana ..... CDC and NCI	Ohio ..... CDC NPCR
Alaska ..... CDC NPCR	Maine ..... CDC NPCR	Oklahoma ..... CDC NPCR
American Samoa ..... CDC NPCR	Marshall Islands ..... CDC NPCR	Oregon ..... CDC NPCR
Arizona ..... CDC NPCR	Maryland ..... CDC NPCR	Palau ..... CDC NPCR
Arkansas ..... CDC NPCR	Massachusetts ..... CDC and NCI	Pennsylvania ..... CDC NPCR
California ..... CDC and NCI	Michigan ..... CDC NPCR	Puerto Rico ..... CDC NPCR
Colorado ..... CDC NPCR	Micronesia ..... CDC NPCR	Rhode Island ..... CDC NPCR
Connecticut ..... NCI SEER	Minnesota ..... CDC NPCR	South Carolina ..... CDC NPCR
Delaware ..... CDC NPCR	Mississippi ..... CDC NPCR	South Dakota ..... CDC NPCR
District of Columbia ..... CDC NPCR	Missouri ..... CDC NPCR	Tennessee ..... CDC NPCR
Florida ..... CDC NPCR	Montana ..... CDC NPCR	Texas ..... CDC and NCI
Georgia ..... CDC and NCI	Nebraska ..... CDC NPCR	U.S. Virgin Islands ..... CDC NPCR
Guam ..... CDC NPCR	Nevada ..... CDC NPCR	Utah ..... CDC and NCI
Hawaii ..... NCI SEER	New Hampshire ..... CDC NPCR	Vermont ..... CDC NPCR
Idaho ..... CDC and NCI	New Jersey ..... CDC and NCI	Virginia ..... CDC NPCR
Illinois ..... CDC and NCI	New Mexico ..... NCI SEER	Washington ..... CDC NPCR
Indiana ..... CDC NPCR	New York ..... CDC and NCI	West Virginia ..... CDC NPCR
Iowa ..... NCI SEER	North Carolina ..... CDC NPCR	Wisconsin ..... CDC NPCR
Kansas ..... CDC NPCR	North Dakota ..... CDC NPCR	Wyoming ..... CDC NPCR
Kentucky ..... CDC and NCI	Northern Marianas ..... CDC NPCR	

How to Access the Data

The data in the U.S. Cancer Statistics public use database come from CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. They are analyzed using SEER*Stat software, which is distributed by the SEER Program.

To obtain access to the U.S. Cancer Statistics database, you must first obtain SEER data Research Plus access and then submit the U.S. Cancer Statistics Database Request Form.

1. SEER offers two data products: SEER Research data and SEER Research Plus data. You must have access to SEER Research Plus data (see https://seer.cancer.gov/data/access.html). This step is not complete until you have completed the user authentication process and received email confirmation that your SEER*Stat account was created or updated.

2. Complete the U.S. Cancer Statistics Database Request Form at www.cdc.gov/united-states-cancer-statistics/public-use/pdf/uscs-public-use-database-request-form-fillable-508.pdf.

3. Email the form to uscsdata@imsweb.com. Your request for access will be processed within 2 business days and you will receive a response notifying you when you have access to the databases in the SEER*Stat software.

4. Download and install the SEER*Stat software at https://seer.cancer.gov/seerstat/. The site also provides training tutorials.

A new SEER data use agreement and U.S. Cancer Statistics Database Request Form are required for each data submission year. You will receive email reminders to complete these forms when new data become available.

Cautionary Notes

Case inclusions and exclusions

Questions?

If you have questions, please contact CDC at uscsdata@cdc.gov.

Cancer registries that are supported by CDC's National Program of Cancer Registries (NPCR) or the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) program report all incident cases coded as *in situ* (non-malignant), invasive (malignant; primary site only), and non-malignant (including borderline and benign) central nervous system tumors according to the *International Classification of Diseases for Oncology, Third Edition* (ICD-O-3), with the following exceptions:

- *In situ* cancers of the cervix are not reported.
- Basal and squamous cell carcinomas of the skin are not reported, except when these occur on the skin of the genital organs.
- *In situ* cancers of the urinary bladder are re-coded as invasive behavior because the information that distinguishes between *in situ* and invasive bladder cancers is not always available or reliable. Stage for these cases remains coded as *in situ*.[1]

Additionally, in this database:

- Cases with an unknown age or sex have been excluded from the database. The frequency counts will not change based on whether *Known Age* or *Male or Female Sex* is checked on the SEER*Stat Selection tab.
- *Malignant Behavior* is a default selection for this database, as this restriction is used by CDC's NPCR and NCI's SEER Program for generating most official cancer statistics. Malignant behavior is defined by the variable *Behavior Code ICD-O-3*. This database includes *in situ* and nonmalignant central nervous system (CNS) cases. These nonmalignant cases can be analyzed by unselecting the *Malignant Behavior* check box on the SEER*Stat Selection tab.

Impact of COVID-19 on cancer incidence data

In March 2020, the World Health Organization declared COVID-19 a pandemic. Soon after, stay-at-home orders, business and school shutdowns, and travel advisories were implemented in the United States to prevent the spread of COVID-19. Additionally, some health care systems reduced access to routine care. These measures interrupted cancer screening, diagnosis, and care as people postponed or deferred health care visits, particularly between March and May 2020.[2]

Impact of COVID-19 on joinpoint trends

The decline in cancer incidence in 2020 was likely an impact of the COVID-19 pandemic; estimates such as cancer incidence trends may be biased as a result. The joinpoint regression model for the analysis of trends was not designed to accommodate a one-year anomaly in data. When using joinpoint regression, inclusion of the 2020 data may influence the location of joinpoints, the value of the trend measure (annual percent change) and provide a poor fit of the model and larger confidence intervals. This may lead to incorrect interpretations of population-level cancer prevention and early detection efforts.

CDC and NCI include the 2020 incidence rates in statistical reports and graphics, but do not include them in joinpoint models. The 2021 incidence data will be included in statistical reports and joinpoint models.[3] JoinPoint software allows

researchers to exclude incidence data for one or more years from trend analyses. Exclude 2020 data for incidence trend analyses, but data for diagnosis year 2021 and later years can be included in incidence trend analyses.

Suppression rules

Suppressing fewer than 16 cases

The suppression rule[4,5] is fewer than 16 cases for the time period based on rate stability. This suppression rule is applied automatically in this database.

When the number of cases used to compute the incidence rates is small, those rates tend to have poor reliability. Therefore, to discourage misinterpretation and misuse of unstable counts, rates, and trends, these statistics are not shown in tables and figures when the counts are fewer than 16 for the time period. Counts of 16 or fewer in a numerator result in a standard error that is at least 25% as large as the rate itself. Equivalently, a count of 16 or fewer 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 close to the observed population age distribution.

Another important reason for employing a cell suppression threshold value is to protect the confidentiality of patients whose data are included in a report by reducing or eliminating the risk of identity disclosure. The cell suppression threshold value of 16 is recommended to protect patient confidentiality given the low level of geographic and clinical detail provided.

Complementary cell suppression

Complementary cell suppression prevents users from unmasking suppressed counts by suppressing at least two cells to prevent the calculation of a third cell. For example, if two categories (male and female) and the total number are reported, both male and female categories must be suppressed. If both categories are not suppressed, the number of males could be subtracted from the total to determine the number of females. Use this suppression when a single year or multiple years of data are presented.

- If a single state in the nation is suppressed, suppress counts for the nation. Rates, confidence intervals, and populations can be shown at the national level.
- If a single state in a region is suppressed, suppress counts for the region and the nation. Rates, confidence intervals, and populations can be shown at the regional and national levels.

Race and ethnicity suppression

States have the option to suppress race-specific and Hispanic ethnicity-specific data every submission year. While these states can be included in an aggregated analysis, the affected state's race and ethnicity information cannot be reported at the state level.

The merged system-supplied variable, *state race ethnicity suppress,* can be used to restrict your analysis to the states that are eligible to be included in a state-level analysis of race and ethnicity combinations. If conducting a state-level analysis of race or ethnicity only, manually make restrictions in the SEER*Stat Selection tab.

The following states have data presentation restrictions:

- Data for Hispanic and non-Hispanic American Indian and Alaska Native people cannot be displayed for Illinois, Kansas, New Jersey, and New York.
- Data for Hispanic Asian and Pacific Islander and Hispanic Black people cannot be displayed for Kansas.

For more information, please refer to the *Race recode (W, B, AIAN, API), Origin recode NHIA (Hispanic, Non-Hisp),* and *Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)* variable descriptions.

Case-level data

As a further mechanism to protect data confidentiality and due to data sharing agreements with some states, the case listing function in SEER*Stat has been disabled for this database.

Nonmalignant (benign and borderline) central nervous system (CNS) tumors

Cancer registries began collecting information on nonmalignant brain and other central nervous system tumors with cases diagnosed in 2004. 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 other central nervous system tumors with a behavior code of 0 (benign) or 1 (borderline), in addition to *in situ* and malignant tumors. Data for nonmalignant brain and other nervous system tumors were available from all registries contributing to this report.

Behavior

The behavior variable in the current database is *Behavior Code ICD-O-3*. Previous database releases included the variable *Behavior Recode for Analysis*.

The database's default is to restrict analyses to malignant cases. CDC's NPCR and NCI's SEER Program use this restriction when generating most official cancer statistics. To analyze benign, borderline, or *in situ* cases, uncheck the "Malignant Behavior" box in the SEER*Stat Selection tab.

To create comparable analyses using a database with data from submission years 2018 and earlier:

- Uncheck the "Malignant Behavior" box in the SEER*Stat Selection tab.
- Add the following selection criteria: {Site and Morphology.Behavior recode for analysis} = 'Malignant','Only malignant in ICD-O-3','Only malignant 2010+'.

Primary site variables

Beginning in diagnosis year 2010, some lymphoma and leukemia ICD-O-3 codes were updated based on changes from the World Health Organization. The appropriate site recode variable to include these updates for cancer cases of all ages is *Site recode ICD-O-3/WHO 2008*. For childhood cancers, the *International Classification of Childhood Cancer (ICCC) recode 3rd edition ICD-O-3/IARC 2017* and *ICCC recode extended 3rd edition ICD-O-3/IARC 2017* variable definitions are included in the database.[6,7,8,9,10]

Consider reviewing the variable Site recode ICD-O-3/WHO 2008 before using the directly coded primary site. See more information on the SEER primary site recodes at https://seer.cancer.gov/siterecode/.

Stage

A merged variable, *Merged Summary Stage,* is provided to span time periods when three different staging schemes are used. The following sections describe the coding logic for this merged variable.

For NPCR registries

- If a case was diagnosed in 2001, 2002, 2003, 2016 or 2017, stage at diagnosis is recorded using the *SEER Summary Stage 2000* variable value.
- If a case was diagnosed in or between 2004 and 2015, stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value. If the *Derived SEER Summary Stage 2000* variable is blank or unstaged, and the *SEER Summary Stage 2000* variable has a valid value, that value is used to populate the merged variable.
- If a case was diagnosed in 2018 or later, stage at diagnosis is recorded using the *Summary Stage 2018* variable value.

For SEER-only registries (Connecticut, Hawaii, Iowa, and New Mexico)

- If a case was diagnosed in 2001, 2002, or 2003, stage at diagnosis is recorded using the *SEER Summary Stage 2000* variable value.
- If a case was diagnosed in or between 2004 and 2017, stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value.
- If a case was diagnosed in 2018 or later, stage at diagnosis is recorded using the *Derived Summary Stage 2018* variable value.

Notes

- Due to changes made in the Summary Stage 2018 Coding Manual, for cases diagnosed in 2018 or later:
  - The category *Regional, NOS* (code 5) is no longer used.
  - There is an artificial increase in the category *Regional by Direct Extension Only* (code 2) for brain, CNS Other, and lymphoma cases. This is because *Regional, NOS* for these cases changed from code 5 to code 2.
- *Merged Summary Stage* data are not available for testis cases.

Reporting delay

NPCR and SEER registries annually submit all eligible years of data to CDC and NCI, respectively. As a result, cases submitted in previous years may be deleted, and new cases diagnosed in previous years may be added. The addition of new cases is called a *reporting delay*. This reporting delay may cause an appearance of decreasing trends.[11] For example, reporting of melanoma cases diagnosed in an outpatient facility may be delayed. As a result, the trend in incident melanoma cases might superficially appear to have dropped in the most recent year.

Checking SEER*Stat frequencies

You can check the setup of your SEER*Stat program by comparing results to those published in the U.S. Cancer Statistics Data Visualizations tool. Note that most of the data in the Data Visualizations tool are restricted to malignant behaviors. Be sure the Malignant Behavior box is selected in the SEER*Stat Selection tab.

References

1. Surveillance, Epidemiology, and End Results Program. SEER Coding and Staging Manual. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2025.
2. IQVIA Institute for Human Data Science. *Shifts in Healthcare Demand, Delivery, and Care During COVID-19 Era*. April 2020.
3. Surveillance, Epidemiology, and End Results Program. Impact of COVID on SEER Data Releases. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2024.
4. Federal Committee on Statistical Methodology. *Report on Statistical Disclosure Limitations Methodology (Statistical Working Paper 22)*. Washington, DC: Office of Management and Budget; 2005.
5. Doyle P, Lane JI, Theeuwes JM, Zayatz LM. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam: Elsevier Science; 2001.
6. Fritz A, Percy C, Jack A, et al., editors. *International Classification of Diseases for Oncology, Third Edition*. Geneva: World Health Organization; 2000.

7. *International Classification of Diseases for Oncology, Third Edition, First Revision.* Geneva: World Health Organization, 2013.

8. Ruhl J, Adamo M, Dickie L., Negoita, S. Hematopoietic and Lymphoid Neoplasm Coding Manual. National Cancer Institute, Bethesda, MD, 2024.

9. Surveillance, Epidemiology, and End Results Program. Solid Tumor Rules. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2025.

10. Surveillance, Epidemiology, and End Results Program. Hematopoietic and Lymphoid Neoplasm Database. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2024.

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

Documentation

The United States Cancer Statistics public use database for 2001–2022:

- Includes race and ethnicity variables.
- Does not include Puerto Rico data, which are available upon special request.
- The population denominators are county-level intercensal estimates (for July 1, 2001–2009, July 1, 2010–2019) and vintage 2023 postcensal estimates (for July 1, 2020–2023) by age, sex, bridged race, and ethnicity, aggregated to the state and national levels.

Population coverage by diagnosis year

- In 2001 and 2002, cases that were diagnosed in Mississippi are not available. The U.S. population covered for 2001 and 2002 is 99.0%.
- For cases diagnosed from 2003 through 2022, 100% of the population is covered for all 50 states and the District of Columbia.
- The U.S. population coverage for 2001 through 2022 is 99.9%.

U.S. and Puerto Rico public use database

A database including U.S. and Puerto Rico data is no longer released. If you would like to analyze data from Puerto Rico, please contact us at uscsdata@cdc.gov. For information on the previously released U.S. and Puerto Rico database, please review the U.S. Cancer Statistics 2005–2020 Public Use Database Data Standards and Data Dictionary (available at www.cdc.gov/united-states-cancer-statistics/public-use/pdf/uscs-public-use-database-technical-documentation-uspr-2005-2020-508.pdf).

Change History

June 2025 release

- Cases diagnosed in Indiana in 2020 and 2021 that had been previously excluded met the USCS publication criteria and were included in this year's database release.
- The definitions of various human papillomavirus (HPV)-associated cancers (anal and rectal squamous cell carcinoma, vulvar squamous cell carcinoma, vaginal squamous cell carcinoma and penile squamous cell carcinoma) were updated to include ICD-0-3-histology codes 8085/3 and 8086/3. These ICD-0-3 histology codes incorporate p16 immunohistochemistry testing results that describe whether HPV is present (8085/3 indicates HPV-positive squamous cell carcinoma) or absent (8086/3 indicates HPV-negative squamous cell carcinoma).
- The definition of acute myeloid leukemia in the *tobacco-related cancer* variable was updated to include ICD-O-3 histology codes 9877/3 (acute myeloid leukemia with mutated NPM1), 9878/3 (acute myeloid leukemia with biallelic mutations of CEBPA), 9879/3 (acute myeloid leukemia with mutated RUNX1), and 9912/3 (acute myeloid leukemia with BCR-ABL1). These changes were made to reflect updated ICD-O-3 codes.

June 2024 release

- The Schema ID variable was added to the public use database for cases diagnosed in 2018 onward. This variable links the site-specific data items (SSDIs) with the appropriate primary site and histology.
- The variables for the classification of childhood cancers were changed from International Classification of Childhood Cancer (ICCC) site recode ICD-O-3/WHO 2008 and ICCC site recode extended ICD-O-3/WHO 2008 to International Classification of Childhood Cancer (ICCC) recode 3rd edition ICD-O-3/IARC 2017 and ICCC recode extended 3rd edition ICD-O-3/IARC 2017. The SEER Program defined these variables based on definitions presented in International Classification of Childhood Cancer, Third Edition and IARC classifications.[1]
- Data for the *Merged estrogen receptor* and *Merged progestogen receptor* variables are available from 2010 onward. Data for the *Merged HER2 summary* variable are available from 2011 onward.
- Suppression is no longer required for data describing Hispanic or non-Hispanic ethnicity for North Dakota and Wisconsin.

June 2023 release

- When data are presented by state, cases for non-Hispanic American Indian/Alaska Native people in Illinois should be suppressed. Suppression is no longer required for other race and ethnicity combinations in Illinois. The *state race eth suppress* variable has been updated to reflect this change.
- Data describing Hispanic or non-Hispanic ethnicity are suppressed for North Dakota and Wisconsin in the *Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic), Origin recode NHIA (Hispanic, Non-Hisp),* and *state race eth suppress* variables.
- Data describing the type of surgery to the primary site performed as part of the first course of treatment (*Rx summary – surgery primary site* variable) are available only for diagnosis years 2010 or later.

June 2022 release

- The variable describing race was revised and renamed to *Race recode (W, B, AIAN, API)*; it was *Race recode for USCS* in previous years. The "other" and "unknown" categories were collapsed to one "Unknown" category.
- The *Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)* variable was added to describe both race and ethnicity.
- Delaware, Kentucky, and Pennsylvania no longer require race and ethnicity suppressions when data are presented by state. The *state race eth suppress* variable has been updated to reflect this change.
- Three changes were made to *Merged Summary Stage:*
    - Testis cases diagnosed in 2018 and 2019 are excluded. In the database released in June 2021, stage data for all testis cases were excluded.
    - Myeloma and leukemia cases are included. In the database released in June 2021, stage data for all myeloma and leukemia cases were excluded.
    - For brain and central nervous system cases, if the behavior was coded as benign or borderline, the variable was set to benign/borderline.

June 2021 release

- To protect confidentiality further while allowing access to all cases, the *Type of Reporting Source* variable is not included in the database. Excluding this variable allows cases identified through death certificates and autopsy reports to be included in the database. This change means that statistics calculated using the public use database now match statistics in the U.S. Cancer Statistics Data Visualizations and CDC WONDER tools.
- Two revised site recode variables were added:
    - *AYA site recode 2020.*
    - *Lymphoid neoplasm recode 2021*.
- The following variables were added:
    - *Grade clinical* (available for cases diagnosed in 2018 or later).
    - *Grade pathological* (available for cases diagnosed in 2018 or later).
    - *Merged estrogen receptor* (available for breast cancer cases diagnosed in 2004 or later). This variable replaces *CS Site Specific Factor 1* for breast cancers.
    - *Merged progesterone receptor* (available for breast cancer cases diagnosed in 2004 or later). This variable replaces *CS Site Specific Factor 2* for breast cancers.
    - *Merged HER2 summary* (available for breast cancer cases diagnosed in 2010 or later). This variable replaces *CS Site Specific Factor 15* for breast cancers.

Reference

1. Steliarova-Foucher E, Colombet M, Ries LAG, Rous B, Stiller CA. Classification of tumours. In: Steliarova-Foucher E, Colombet M, Ries LAG, et al. *International Incidence of Childhood Cancer, Volume III.* Lyon: International Agency for Research on Cancer, In press.

Analyses Checklist

Multi-year analyses

The database includes variables that can be used to restrict analyses to the states meeting U.S. Cancer Statistics publication criteria during the most commonly analyzed multi-year time periods, specifically:

- All years of data in the database (variable *USCS0122* for diagnosis years 2001–2022).
- The most recent 10 years of data (*USCS1322* for diagnosis years 2013–2022).
- The most recent 5 years of data (*USCS1822* for diagnosis years 2018–2022).

If you are conducting a multi-year analysis and want to restrict it to the states that met publication criteria during each of the years, did you use variable *USCS0122, USCS1322,* or *USCS1822* and also use the *Year of Diagnosis* variable on the SEER*Stat Selection tab?

- This is important for trend analyses so the same states are included for each year.
- The *Year of Diagnosis* variable is used in combination with the predefined USCS variable to exclude the non-relevant years. For example, if *USCS1822* is used, then *Year of Diagnosis* should also be restricted to diagnosis years 2018–2022 in the SEER*Stat Selection tab.
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at uscsdata@cdc.gov and we will create a new variable for you.

Single-year analyses

If you are analyzing just 1 year of data, did you use the variable *USCS Standard* and restrict the analysis to the specific Year of Diagnosis in the SEER*Stat Selection tab?

Common selection and reporting considerations

State-level race, ethnicity, or race/ethnicity combinations

If you are reporting state-level race, ethnicity, or race/ethnicity combinations, have you suppressed data from the registries that opted out of reporting these data items? Race and ethnicity combinations can be excluded using the *State Race Ethnicity Suppress* variable. Race-only or ethnicity-only suppressions should be done manually in the SEER*Stat Selection tab.

User-defined primary site variable

If a user-defined primary site variable was created (rather than using the *Site recode ICD-O-3/WHO 2008* variable):

- Did you exclude leukemias and lymphomas (9590–9992)?
- Did you consider excluding Kaposi sarcoma (9140) and mesothelioma (9050–9055)?

For more information, see the Primary Site Variables description.

Histology

If your analysis includes histology, and if appropriate for the cancer site, did you use the *Diagnostic Confirmation* variable to specify the analysis be limited to microscopically confirmed cases?

Sex-specific cancers

If you are analyzing sex-specific cancers such as prostate cancer or female breast cancer, did you limit the analysis to the appropriate sex to get the correct population denominator?

Rates

When reporting rates, have you included the label "per 100,000 persons," "per 100,000 women," or "per 100,000 men"?

Citations

Have you included citations for the:

- Percentage of United States population coverage provided by the database?
- U.S. Cancer Statistics 2001–2022 Public Use Research Database?

Variable Definitions: Age at Diagnosis

Age recode with <1 year olds

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from Age at Diagnosis
Source item number: 230

This variable indicates age range at diagnosis by grouping patient into one of 19 categories (0, 1–4, 5–9, …, 75–79, 80–84, ≥85 years). Derived from the NAACCR variable *Age at Diagnosis [230],* which is the age (in years) of the patient at diagnosis.

Considerations for use

Different primary tumors for the same patient may have different values. Records for persons with multiple primary cancers cannot be identified in this database.

Values

- 00 years
- 01–04 years
- 05–09 years
- 10–14 years
- 15–19 years
- 20–24 years
- 25–29 years
- 30–34 years
- 35–39 years
- 40–44 years
- 45–49 years
- 50–54 years
- 55–59 years
- 60–64 years
- 65–69 years
- 70–74 years
- 75–79 years
- 80–84 years
- ≥85 years

Variable Definitions: Race, Sex, Year of Diagnosis, and Registry

Sex

Source of standard: North American Association of Central Cancer Registries
Source item name: Sex
Source item number: 220

This variable indicates the sex of the patient.

Considerations for use

- To get the correct population denominator, "female" must be selected when analyzing female-specific cancers (such as ovarian cancer or female breast cancer) and "male" for male-specific cancers (such as prostate cancer).
- Due to small case counts, cases for unknown sex are excluded from this database.

Values

- Male
- Female

Year of diagnosis

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from *Date of initial diagnosis (CoC)*
Source item number: 390

This variable indicates the year of initial diagnosis by a recognized medical practitioner for the cancer being reported, whether clinically or microscopically confirmed. Derived from *Date of initial diagnosis (CoC) [390]*.

Considerations for use

As an additional confidentiality measure, date of diagnosis is not provided.

More information

- NAACCR Data Standards and Data Dictionary
- SEER Race Recode Changes (https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/)
- Standards for Oncology Registry Entry (STORE)
- SEER Program Coding and Staging Manual (https://seer.cancer.gov/tools/codingmanuals/)

Values

- 2001
- ...
- 2022

Addr at DX – state

Source of standard: North American Association of Central Cancer Registries
Source item name: *Addr at DX–State*
Source item number: 80

This variable indicates the U.S. state in which the patient lived at the time the reportable tumor was diagnosed.

Considerations for use

- If a patient has multiple tumors, the state of residence at the time of diagnosis may differ for each.
- Multiple records for an individual with more than one primary cancer cannot be linked in this database.

More information

- NAACCR Data Standards and Data Dictionary
- Facility Oncology Registry Data Standards (FORDS) variable *State at Diagnosis*

Values

- Alaska
- Alabama
- Arkansas
- Arizona
- California
- Colorado
- Connecticut
- District of Columbia
- Delaware
- Florida
- Georgia
- Hawaii
- Idaho
- Illinois
- Indiana
- Iowa
- Kansas
- Kentucky
- Louisiana
- Massachusetts
- Maryland
- Maine
- Michigan
- Minnesota
- Missouri
- Mississippi
- Montana
- North Carolina
- North Dakota
- Nebraska
- New Hampshire
- New Jersey
- New Mexico
- Nevada
- New York
- Ohio
- Oklahoma
- Oregon
- Pennsylvania
- Rhode Island
- South Carolina
- South Dakota
- Tennessee
- Texas
- Utah
- Virginia
- Vermont
- Washington
- Wisconsin
- West Virginia
- Wyoming

USCS standard

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from *Addr at DX–State* and U.S. Cancer Statistics publication criteria
Source item number: Derived from NAACCR's 80

This variable indicates a central cancer registry met the U.S. Cancer Statistics publication criteria for a single year of analysis.

Considerations for use

- This variable allows the selection of only cancer registries with data that meet the U.S. Cancer Statistics publication criteria for an individual diagnosis year. Specify the year of diagnosis in the SEER*Stat Selection tab.

- If you are conducting a multiyear analysis and want to restrict the analysis to the registries that met publication criteria for each of those years (for example, a trend analysis), we recommend using the predefined variables *USCS1822* (includes diagnosis years 2018–2022), *USCS1322* (includes diagnosis years 2013–2022), or *USCS0122* (includes diagnosis years 2001–2022).

- If you would like to analyze a range of years other than those predefined variables, please contact CDC at uscsdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.

Values

Number of central cancer registries	Year of diagnosis
50	2001 and 2002
51	2003–2022

Race recode (W, B, AIAN, API)

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from *Race 1* and *IHS Link*
Source item number: 160 *(Race 1)* and 192 *(IHS Link)*

This variable indicates the derived code for the patient's race. Race is coded separately from Hispanic ethnicity.

Data quality checks code a non-White race before a White race. This variable is created using NAACCR variables *Race 1* and the Indian Health Service (IHS) link. If *Race 1* is White and there is a positive *IHS Link,* then *Race/Ethnicity* is set to American Indian/Alaska Native (AI/AN).

Considerations for use

States have the option to suppress race-specific and Hispanic-specific data every submission year. While these states can be included in an aggregated analysis, their race and ethnicity information cannot be reported at the state level. The following states have state-level race data presentation restrictions:

- Data for Hispanic and non-Hispanic American Indian and Alaska Native people cannot be displayed for Illinois, Kansas, New Jersey, and New York.
- Data for Hispanic Asian and Pacific Islander and Hispanic Black people cannot be displayed for Kansas.

Race is defined by specific physical, hereditary, and cultural traditions or origins, not necessarily by birthplace, place of residence, or citizenship. "Origin" is defined by the U.S. Census Bureau as the heritage, nationality group, lineage, or in some cases, the country of birth of the person or the person's parents or ancestors before their arrival in the United States. As a standard practice, central cancer registries classify race as coded in the medical record. To address AI/AN misclassification in cancer registry data, registries supported by CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program link their central cancer registry data to the Indian Health Service (IHS) administrative records database.

- SEER registries link their data annually, with the most recent linkage occurring among cases diagnosed from 1998 to 2022. Annually, 32 NPCR registries with at least one Purchase/Referred Care Service Delivery Area (PRCSDA) county in their state link their data. All NPCR registries link every 5 years, with the most recent linkage occurring in 2021.
- Although the linkage with IHS does not completely resolve the classification of race for AI/AN cases, it helps provide a more comprehensive and accurate picture of the cancer burden in this population.
- When interpreting data results, be aware that AI/AN populations may be undercounted during years when linkages did not occur in all NPCR registries.
- If a project is looking specifically at AI/AN populations, analysts may consider using the USCS Data Visualizations AI/AN restricted to PRCDA only module.

In all separate records of tumors for the same patient, the patient has the same race code.

This variable contains "other unspecified" and "unknown" categories. These groups are coded as "unknown race" for the purpose of analyses as specified in the SEER documentation at https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/. Population data are not available for the "other race" and "unknown race" categories.

NPCR–Indian Health Service (IHS) linkage schedule

All NPCR-funded registries link with the IHS every 5 years. The most recent linkage year was 2021.

All state central cancer registries with at least one Purchase/Referred Care Delivery Area (PRCDA) county (previously referred to as Contract Health Service Delivery Area [CHSDA] counties) link with the IHS every year. These include: Alabama, Alaska, Arizona, California, Colorado, Florida, Idaho, Indiana, Kansas, Louisiana, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Montana, Nebraska, Nevada, New York, North Carolina, North Dakota, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Texas, Utah, Virginia, Washington, Wisconsin, and Wyoming.

Values

- All Races
- White
- Black
- American Indian/Alaska Native
- Asian or Pacific Islander
- Unknown (including other unspecified 1991+)

More information

- NAACCR Data Standards and Data Dictionary
- SEER Race Recode Changes (https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/)
- Standards for Oncology Registry Entry (STORE)
- SEER Program Coding and Staging Manual (https://seer.cancer.gov/tools/codingmanuals/)

Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from *Race 1, IHS Link,* and *Origin recode NHIA (Hispanic, Non-Hisp)*
Source item numbers: 160 *(Race 1),* 192 *(IHS Link),* and 191 *(NHIA Derived Hisp Origin)*

This variable indicates the derived code for the patient's race and Hispanic ethnicity. It is obtained by merging the race variable, *Race recode (W, B, AIAN, API)* and Hispanic ethnicity, *Origin recode NHIA (Hispanic, Non-Hisp)* variables.

Considerations for use

States have the option to suppress race-specific and Hispanic-specific data every submission year. While these states can be included in an aggregated analysis, their race and ethnicity information cannot be reported at the state level. The following states have state-level race data presentation restrictions:

- Data for Hispanic and non-Hispanic American Indian and Alaska Native people cannot be displayed for Illinois, Kansas, New Jersey, and New York.

- Data for Hispanic Asian and Pacific Islander and Hispanic Black people cannot be displayed for Kansas.

Race is defined by specific physical, hereditary, and cultural traditions or origins, not necessarily by birthplace, place of residence, or citizenship. "Origin" is defined by the U.S. Census Bureau as the heritage, nationality group, lineage, or in some cases, the country of birth of the person or the person's parents or ancestors before their arrival in the United States. As a standard practice, central cancer registries classify race as coded in the medical record. To address AI/AN misclassification in cancer registry data, registries supported by CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program link their central cancer registry data to the Indian Health Service (IHS) administrative records database.

- SEER registries link their data annually, with the most recent linkage occurring among cases diagnosed from 1998 to 2022. Annually, 32 NPCR registries with at least one Purchase/Referred Care Service Delivery Area (PRCSDA) county in their state link their data. All NPCR registries link every 5 years, with the most recent linkage occurring in 2021.

- Although the linkage with IHS does not completely resolve the classification of race for AI/AN cases, it helps provide a more comprehensive and accurate picture of the cancer burden in this population.
- When interpreting data results, be aware that AI/AN populations may be undercounted during years when linkages did not occur in all NPCR registries.
- If a project is looking specifically at AI/AN populations, analysts may consider using the USCS Data Visualization Tools AI/AN restricted to PRCDA only module.

In all separate records of tumors for the same patient, the patient has the same race code.

This variable contains "other unspecified" and "unknown" categories. These groups are coded as "unknown race" for the purpose of analyses as specified in the SEER documentation at https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/. Population data are not available for the "other race" and "unknown race" categories.

NPCR–Indian Health Service (IHS) linkage schedule

All NPCR-funded registries link with the IHS every 5 years. The most recent linkage year was 2021.

All state central cancer registries with at least one Purchase/Referred Care Delivery Area (PRCDA) county (previously referred to as Contract Health Service Delivery Area [CHSDA] counties) link with the IHS every year. These include: Alabama, Alaska, Arizona, California, Colorado, Florida, Idaho, Indiana, Kansas, Louisiana, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Montana, Nebraska, Nevada, New York, North Carolina, North Dakota, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Texas, Utah, Virginia, Washington, Wisconsin, and Wyoming.

Values

- Non-Hispanic White
- Non-Hispanic Black
- Non-Hispanic American Indian/Alaska Native
- Non-Hispanic Asian or Pacific Islander
- Hispanic (All Races)
- Non-Hispanic Unknown Race

More information

- NAACCR Data Standards and Data Dictionary
- SEER Race Recode Changes (https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/)
- Standards for Oncology Registry Entry (STORE)
- SEER Program Coding and Staging Manual (https://seer.cancer.gov/tools/codingmanuals/)

Program

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Addr at DX–State* and U.S. Cancer Statistics publication criteria
Source item number: Derived from NAACCR's 80

This variable indicates whether a state is funded by CDC's NPCR or NCI's SEER Program.

Considerations for use

- Central cancer registries that received funding from CDC and submitted data during any 2001–2022 diagnosis year are categorized as "NPCR registries." They include Alabama, Alaska, Arizona, Arkansas, California, Colorado, Delaware, District of Columbia, Florida, Georgia, Idaho, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode

Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming.

- "SEER registries" are central cancer registries receiving funding only from NCI during the 2001–2022 diagnosis years (Connecticut, Hawaii, Iowa, and New Mexico).

Values

- NPCR
- SEER

Region

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Addr at DX–State* and U.S. Census region
Source item number: Derived from NAACCR's 80

This variable indicates the U.S. Census region in which the patient lived at the time of diagnosis. The NAACCR data item *Address at Diagnosis-State* is recoded into one of the four U.S. Census regions: Northeast, Midwest, South, and West.

Considerations for use

- There is a potential for bias in the incidence rates for Census regions, as only data from registries that met U.S. Cancer Statistics publication criteria are included in the database. It is encouraged that age-adjusted incidence rates for U.S. Census regions be presented only if:
    - At least 80% of the population for the Census region was covered by cancer registries that met U.S. Cancer Statistics publication criteria.
    - The 95% confidence intervals around the observed age-adjusted regional 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, and female lung and bronchus) included the estimate of the regional rate calculated.

- If any state in a region has a case count of fewer than 16, then the case counts for U.S. Census regions cannot be presented. See the Census Geographic Areas Reference Manual, Chapter 6: Statistical Groupings of States and Counties at https://www2.census.gov/geo/pdfs/reference/GARM/Ch6GARM.pdf for a list of states in each region.

Values

- Northeast
- Midwest
- South
- West

USCS0122

Source of standard: CDC's National Program of Cancer Registries
Source item name: Not applicable
Source item number: Not applicable

This variable indicates whether a central cancer registry met the U.S. Cancer Statistics publication criteria for all cancer sites combined each year in 2001–2022.

Considerations for use

- This variable is used for analysis of combined 2001–2022 data in the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2022 database. Data from registries that did not meet the U.S. Cancer Statistics publication criteria in any given year were excluded from the database for that year.
- If you are conducting a multiyear analysis and want to restrict the analysis to the registries that met publication criteria for each of those years (for example, a trend analysis), use the predefined variables *USCS1822* (includes diagnosis years 2018–2022), *USCS1322* (includes diagnosis years 2013–2022), or this variable (includes diagnosis years 2001–2022).
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at uscsdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.

Values

- Does not meet USCS standard 0122
- Meets USCS standard 0122

USCS1322

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Addr at DX–State* and U.S. Cancer Statistics publication criteria
Source item number: Derived from NAACCR's 80

This variable indicates whether a central cancer registry met the U.S. Cancer Statistics publication criteria for all cancer sites combined each year in 2013–2022. When using this variable, restrict the diagnosis years to 2013–2022. This is done in SEER*Stat on the Selection tab using the *year of diagnosis* variable.

Considerations for use

- This variable is used for analysis of combined 2013–2022 data in the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2022 database. Data from registries that did not meet the U.S. Cancer Statistics publication criteria in any given year were excluded from the database for that year.
- If you are conducting a multiyear analysis and want to restrict the analysis to the registries that met publication criteria for each of those years (for example, a trend analysis), use the predefined variables *USCS1822* (includes diagnosis years 2018–2022), this variable (includes diagnosis years 2013–2022), or *USCS0122* (includes diagnosis years 2001–2022).
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at uscsdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.

Values

- Does not meet USCS standard 1322
- Meets USCS standard 1322

USCS1822

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Addr at DX–State* and U.S. Cancer Statistics publication criteria
Source item number: Derived from >NAACCR's 80

This variable indicates whether a central cancer registry met the U.S. Cancer Statistics publication criteria for all cancer sites combined each year in 2018–2022. When using this variable, restrict the diagnosis years to 2018–2022. This is done in SEER*Stat on the Selection tab using the *year of diagnosis* variable.

Considerations for use

- This variable is used for analysis of combined 2018–2022 data in the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2022 database. Data from registries that did not meet the U.S. Cancer Statistics publication criteria in any given year were excluded from the database for that year.
- If you are conducting a multiyear analysis and want to restrict the analysis to the registries that met publication criteria for each of those years (for example, a trend analysis), use this predefined variable (includes diagnosis years 2018–2022), *USCS1322* (includes diagnosis years 2013–2022), or *USCS0122* (includes diagnosis years 2001–2022).
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at uscsdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.

Values

- Does not meet USCS standard 1822
- Meets USCS standard 1822

Origin recode NHIA (Hispanic, non-Hisp)

Source of standard: North American Association of Central Cancer Registries
Source item name: *NHIA Derived Hisp Origin*
Source item number: 191

This variable was derived from the NAACCR standard variables *Spanish/Hispanic Origin [190], Name-Last [2230], Name-Maiden [2390], Birthplace [250], Race 1 [160], IHS Link [192],* and *Sex [220]*.

The NAACCR Hispanic Identification Algorithm (NHIA) uses the combination of these variables to classify cases directly or indirectly as Hispanic for analytic purposes.

Considerations for use

States have the option to suppress race-specific and Hispanic-specific data every submission year. While these states can be included in an aggregated analysis, the race and ethnicity information cannot be reported at the state level. The following states have state-level race or ethnicity data presentation restrictions:

- Data for Hispanic and non-Hispanic American Indian and Alaska Native people cannot be displayed for Illinois, Kansas, New Jersey, and New York.
- Data for Hispanic and non-Hispanic Asian and Pacific Islander people cannot be displayed for Kansas.

Blank values are allowed for states that chose not to include data for NHIA in this file.

Values

- Non-Spanish–Hispanic–Latino
- Spanish–Hispanic–Latino

More information

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

Variable Definitions: Site and Morphology

Primary site – labeled

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from *Primary Site*
Source item number: 400

This variable indicates the topography code from *The International Classification of Diseases for Oncology, Third Edition* (ICD-O-3) for the primary site of the tumor being reported.

Considerations for use

Beginning in diagnosis year 2010, there were updates to some of the lymphoma and leukemia codes. To include these updates, the appropriate primary site variables to use are *Site recode ICD-O-3/WHO 2008* for all ages, and *ICCC site recode ICD-O-3/WHO 2008* for the childhood cancer recodes.

Values

- C00.0–C77.9
- C80.9–Unknown primary site

More information

SEER Program Coding and Staging Manual (https://seer.cancer.gov/tools/codingmanuals/)

Histologic type ICD-O-3 (International Classification of Diseases for Oncology, Third Edition)

Source of standard: North American Association of Central Cancer Registries
Source item name: *Histologic Type ICD-O-3*
Source item number: 522

This variable indicates the morphology code from *The International Classification of Diseases for Oncology,* Third Edition (ICD-O-3) that describes the histologic type (the microscopic composition of cells and tissue for a specific primary tumor) of the primary tumor being reported.

Considerations for use

- This data item is required for cancer cases diagnosed on or after January 1, 2001.
- The histology codes for some tumors may be based on clinical diagnoses, not pathologic confirmation. When analyzing a specific histology, we suggest using the *Diagnostic confirmation* variable in conjunction with this variable. Beginning with 2010 diagnoses, this item also includes histology codes as specified in the *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (2008),* which are listed on pages 4–6 of the NAACCR 2010 Implementation Guidelines.

Values

- 8000–9992

More information

- SEER 2007 Multiple Primary and Histology Coding Rules (https://seer.cancer.gov/tools/mphrules/mphrules_instructions.pdf)

- SEER Hematopoietic Project (https://seer.cancer.gov/tools/heme/)
- ICD-O-3 SEER Site/Histology validation list (https://seer.cancer.gov/icd-o-3/)
- SEER Hematopoietic and Lymphoid Neoplasm Database (https://seer.cancer.gov/seertools/hemelymph)
- SEER Hematopoietic and Lymphoid Neoplasm Coding Manual (https://seer.cancer.gov/tools/heme/Hematopoietic_Instructions_and_Rules.pdf)
- *International Classification of Diseases for Oncology, Third Edition*

Behavior code ICD-O-3

Source of standard: North American Association of Central Cancer Registries
Source item name: *Behavior code ICD-O-3*
Source item numbers: 523

This variable is the code for the behavior of the tumor being reported using ICD-O-3. NAACCR adopted ICD-O-3 as the standard coding system for tumors diagnosed beginning January 1, 2001.

Considerations for use

- This database includes cases with invasive (malignant) and *in situ* benign/borderline behavior for brain and central nervous system cases diagnosed in 2004 and later.
- In SEER*Stat's Selection Tab, the "Malignant Behavior" check box corresponds with Behavior Code ICD-O-3=3 (malignant). This is the default selection in this database. This restriction is used by CDC's NPCR and NCI's SEER Program for generating most official cancer statistics. To analyze benign, borderline, or *in situ* cases, uncheck the "Malignant Behavior" box.

Values

- Benign
- Borderline malignancy
- *In situ*
- Malignant

More information

ICD-O-3 Coding Materials (https://seer.cancer.gov/icd-o-3)

Grade

Source of standard: North American Association of Central Cancer Registries
Source item name: *Grade*
Source item number: 440

This variable indicates the grade or degree of differentiation of the primary tumor being reported. The categories are Well differentiated; Grade I, Moderately differentiated; Grade II, Poorly differentiated; Grade III or Undifferentiated; anaplastic; Grade IV.

For lymphomas and leukemias, this field also is used to indicate T-, B-, Null-, or NK-cell origin.

Considerations for use

- Data are available for cases diagnosed in 2001 through 2017. This data item was not collected for cases diagnosed in 2018 or later.
- The practice of grading varies greatly among pathologists throughout the world, and many malignant tumors are not graded routinely. Since different grading systems may be used, review the SEER site-specific modules at

https://training.seer.cancer.gov/modules_site_spec.html and the FORDS manual that corresponds with the diagnosis year. Each module has an abstracting, coding, and staging section, which has a morphology and grading subsection. Some modules, but not all, contain notes about the grading system that may have been used. Currently, there is no variable to differentiate a specific grading system from another one if more than two grading systems are mentioned.

- Diagnostic practices also influence coding practices. For example, preliminary analysis of tumor grade for prostate cancer showed an increase in the proportion of higher grades from 2002 to 2003. Additional review showed this increase to be artificial, as the International Society of Urologic Pathologists, in conjunction with the World Health Organization (WHO), had made a series of recommendations for modification of the Gleason grading system to reflect contemporary knowledge, alleviate uncertainty, and promote uniformity in its application. One recommendation was for pathologists to report all higher tertiary grade components of the tumor as part of the Gleason score. Another recommendation was for reporting of any higher-grade cancer, no matter how small quantitatively.

- The percentage of cases with a known grade varies by primary cancer site. Rules for coding the tumor grade differ for some primary sites. As a result, it may be appropriate to have a tumor grade coded as "9 – unknown."

- For brain tumor cases diagnosed from 2011 to 2017, cancer registries were required to report the WHO grade classification. Please see the variable description *CS site-specific factor 1* for more information on this brain-specific grade classification.

Values

- Well differentiated; Grade I
- Moderately differentiated; Grade II
- Poorly differentiated; Grade III
- Undifferentiated; anaplastic; Grade IV
- T-cell
- B-cell; pre-B; B-precursor
- Null cell; non T-non B
- NK cell; natural killer cell (1995+)
- Unknown

Grade clinical

Restricted to diagnosis year 2018 or later

Source of standard: North American Association of Central Cancer Registries
Source item name: *Grade Clinical*
Source item number: 3843

This data item records the grade of a solid primary tumor before any treatment, including surgical resection or neoadjuvant.

For cases diagnosed on January 1, 2018, or later, this data item, along with *Grade Pathological,* replaces the data item *Grade* as well as site-specific factors for cancer sites with alternative grading systems such as breast (Bloom-Richardson) and prostate (Gleason).

Considerations for use

- Data are available for cases diagnosed in 2018 and later. This data item was not collected for cases diagnosed in 2001 to 2017.

- For cases that are eligible for American Joint Committee on Cancer (AJCC) staging, the recommended grading system is specified in the AJCC Cancer Staging Manual chapter. The AJCC chapter-specific grading systems (codes

1–5) take priority over the generic grade definitions (codes A–E, L, H, and 9). For cases that are not eligible for AJCC staging, if the recommended grading system is not documented, the generic grade definitions apply.

- Refer to the Site-Specific Data Item (SSDI) Manual and Grade Manual that corresponds with the year when the cases were diagnosed for additional site-specific instructions.

Values

- 1
- 2
- 3
- 4
- 5
- 8
- 9
- A
- B
- C
- D
- E
- L
- H
- M
- S
- Blank(s)

Grade pathological

Restricted to diagnosis year 2018 or later

Source of standard: North American Association of Central Cancer Registries
Source item name: *Grade Pathological*
Source item number: 3844

This data item records the grade of a solid primary tumor that has been resected and for which no neoadjuvant therapy was administered. The highest grade documented from any microscopic specimen of the primary site, whether from the clinical workup or the surgical resection, will be recorded.

For cases diagnosed on January 1, 2018, or later, this data item, along with *Grade Clinical,* replaces the data item *Grade* as well as site-specific factors for cancer sites with alternative grading systems such as breast (Bloom-Richardson) and prostate (Gleason).

Considerations for use

- Data are available for cases diagnosed in 2018 and later. This data item was not collected for cases diagnosed in 2001 to 2017.

- For cases that are eligible for American Joint Committee on Cancer (AJCC) staging, the recommended grading system is specified in the AJCC Cancer Staging Manual chapter. The AJCC chapter-specific grading systems (codes 1–5) take priority over the generic grade definitions (codes A–E, L, H, and 9). For cases that are not eligible for AJCC staging, if the recommended grading system is not documented, the generic grade definitions apply.

- Refer to the Site-Specific Data Item (SSDI) Manual and Grade Manual that corresponds with the year when the cases were diagnosed for additional site-specific instructions.

Values

- 1
- 2
- 3
- 4
- 5
- 8
- 9
- A
- B
- C
- D
- E
- L
- H
- M
- S
- Blank(s)

Diagnostic confirmation

Source of standard: North American Association of Central Cancer Registries
Source item name: *Diagnostic confirmation*
Source item number: 490

This variable records the best method of diagnostic confirmation of the cancer being reported at any time in the patient's history. The rules for coding differ between solid tumors and hematopoietic and lymphoid neoplasms.

Considerations for use

- For analyses that include histology, use the following selection statement in the SEER*Stat Selection tab: "Diagnostic confirmation is = to Microscopically confirmed".

- Diagnostic confirmation is useful to calculate rates based on microscopically confirmed cancers. Complete incidence calculations also include cases that are only confirmed clinically. The percentage of cases that are "clinically diagnosed only" is an indication of whether case finding includes sources outside of pathology reports.

- The microscopically confirmed method has the highest priority for diagnostic confirmation. The remaining values were assigned when the presence of cancer was confirmed with multiple diagnostic methods.

- "Positive histology AND immunophenotyping AND/OR positive genetic studies" (used only for hematopoietic and lymphoid neoplasms M-9590/3-9992/3) was adopted for use beginning with 2010 diagnoses.

Values

- Microscopically confirmed
- Positive histology
- Positive exfoliative cytology, no positive histology
- Pos hist AND immunophenotyping AND/OR pos genetic studies
- Positive microscopic confirm, method not specified
- Positive laboratory test/marker study
- Direct visualization without microscopic confirmation
- Radiography without microscopic confirm
- Clinical diagnosis only
- Unknown

ICD-O-3 histology/behavior, labeled

Source of standard: SEER*Stat recode
Source item name: *ICD-O-3 Hist/behavior, labeled*
Source item number: Not applicable

This variable indicates each ICD-O-3 histology code and behavior code and the respective name of that histology and behavior.

Considerations for use

- This variable is a five-digit ICD-O-3 morphology code. The first four digits indicate the histology (cell type), and the fifth digit is the behavior code.
- Please note that the ICD-O-3 morphology codes have been grouped by major morphology headings as found in the *International Classification of Diseases for Oncology, Third Edition* in the table shown below. However, the morphology codes are not grouped in the database.

Values

- 800: Neoplasms, NOS
- 801–804: Epithelial Neoplasms, NOS
- 805–808: Squamous Cell Neoplasms
- 809–811: Basal Cell Neoplasms
- 812–813: Transitional Cell Papillomas and Carcinomas
- 814–838: Adneomas and Adenocarcinomas
- 839–842: Adnexal and Skin Appendage Neoplasms
- 843: Mucoepidermoid Neoplasms
- 844–849: Cystic, Mucinous and Serous Neoplasms
- 850–854: Ductal and Lobular Neoplasms
- 855: Acinar Cell Neoplasms
- 856–857: Complex Epithelial Neoplasms
- 858: Thymic Epithelial Neoplasms
- 859–867: Specialized Gonadal Neoplasms
- 868–871: Paragangliomas and Glomus Tumors
- 872–879: Nevi and Melanomas
- 880: Soft Tissue Tumors and Sarcomas, NOS
- 881–883: Fibromatous Neoplasms
- 884: Myxomatous Neoplasms
- 885–888: Lipomatous Neoplasms
- 889–892: Myomatous Neoplasms
- 893–899: Complex Mixed and Stromal Neoplasms
- 900–903: Fibroepithelial Neoplasms
- 904: Synovial–Like Neoplasms
- 905: Mesothelial Neoplasms
- 906–909: Germ Cell Neoplasms
- 910: Trophoblastic Neoplasms
- 911: Mesonephromas
- 912–916: Blood Vessel Tumors
- 917: Lymphatic Vessel Tumors
- 918–924: Osseous and Chondromatous Neoplasms
- 925: Giant Cell Tumors
- 926: Miscellaneous Bone Tumors
- 927–934: Odotogenic Tumors
- 935–937: Miscellaneous Tumors
- 938–948: Gliomas
- 949–952: Neuroepitheliomatous Neoplasms
- 953: Meningiomas
- 954–957: Nerve Sheath Tumors
- 958: Granular Cell Tumors and Alveolar Soft Part Sarcomas
- 959–972: Hodgkin and Non–Hodgkin Lymphomas
- 973: Plasma Cell Tumors
- 974: Mast Cell Tumors
- 975: Neoplasms of Histiocytes and Accessory Lymphoid Cells
- 976: Immunoproliferative Disease
- 980–994: Leukemias
- 995–996: Chronic Myeloproliferative Disorders
- 997: Other Hematologic Disorders
- 998–999: Myelodysplastic Syndromes

More information

- *International Classification of Diseases for Oncology, Third Edition, First Revision*
- SEER ICD-O-3 Coding Materials (https://seer.cancer.gov/icd-o-3/)

Laterality

Source of standard: SEER*Stat recode
Source item name: *ICD-O-3 Hist/behavior, labeled*
Source item number: Not applicable

This variable identifies the side of a paired organ, or the side of the body on which the reportable tumor originated. This applies to the primary site only.

Values

- Not a paired site
- Right - origin of primary
- Left - origin of primary
- Only one side - side unspecified
- Bilateral, single primary
- Paired site: midline tumor
- Paired site, but no information concerning laterality

More information

- *Standards for Oncology Registry Entry (STORE)*
- SEER Program Coding and Staging Manual (https://seer.cancer.gov/tools/codingmanuals/)

Site recode ICD-O-3/WHO 2008

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from *Primary Site* and *Histologic Code ICD-O-3*
Source item numbers: 400 *(Primary Site)* and 522 *(Histologic Code ICD-O-3)*

The variable is defined by the SEER Program. Its values are based on NAACCR variables for ICD-O-3, the primary site and histology code of the primary tumor being reported, with updated information for hematopoietic codes based on *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (2008).* The site recode variables define the major cancer sites commonly used in reporting cancer incidence data.

Considerations for use

This is the recommended variable for analyses by primary cancer site.

Values

- All Sites
- Oral Cavity and Pharynx
- Lip
- Tongue
- Salivary Gland
- Floor of Mouth
- Gum and Other Mouth
- Nasopharynx
- Tonsil
- Oropharynx
- Hypopharynx
- Other Oral Cavity and Pharynx
- Digestive System
- Esophagus
- Gallbladder
- Other Biliary
- Pancreas
- Retroperitoneum
- Peritoneum, Omentum and Mesentery
- Other Digestive Organs
- Respiratory System
- Nose, Nasal Cavity and Middle Ear
- Larynx
- Lung and Bronchus
- Pleura
- Trachea, Mediastinum and Other Respiratory Organs
- Kidney and Renal Pelvis
- Ureter
- Other Urinary Organs
- Eye and Orbit
- Brain and Other Nervous System
- Brain
- Cranial Nerves Other Nervous System
- Endocrine System
- Thyroid
- Other Endocrine including Thymus
- Lymphoma
- Hodgkin Lymphoma
- Hodgkin - Nodal

- Stomach
- Small Intestine
- Colon and Rectum
- Colon excluding Rectum
- Cecum
- Appendix
- Ascending Colon
- Hepatic Flexure
- Transverse Colon
- Splenic Flexure
- Descending Colon
- Sigmoid Colon
- Large Intestine, NOS
- Rectum and Rectosigmoid Junction
- Rectosigmoid Junction
- Rectum
- Anus, Anal Canal and Anorectum
- Liver and Intrahepatic Bile Duct
- Liver
- Intrahepatic Bile Duct
- Bones and Joints
- Soft Tissue including Heart
- Skin excluding Basal and Squamous
- Melanoma of the Skin
- Other Non-Epithelial Skin
- Breast
- Female Genital System
- Cervix Uteri
- Corpus and Uterus, NOS
- Corpus Uteri
- Uterus, NOS
- Ovary
- Vagina
- Vulva
- Other Female Genital Organs
- Male Genital System
- Prostate
- Testis
- Penis
- Other Male Genital Organs
- Urinary System
- Urinary Bladder
- Hodgkin - Extranodal
- Non-Hodgkin Lymphoma
- NHL - Nodal
- NHL - Extranodal
- Myeloma
- Leukemia
- Lymphocytic Leukemia
- Acute Lymphocytic Leukemia
- Chronic Lymphocytic Leukemia
- Other Lymphocytic Leukemia
- Myeloid and Monocytic Leukemia
- Acute Myeloid Leukemia
- Acute Monocytic Leukemia
- Chronic Myeloid Leukemia
- Other Myeloid/Monocytic Leukemia
- Other Leukemia
- Other Acute Leukemia
- Aleukemic, Subleukemic and NOS
- Mesothelioma
- Kaposi Sarcoma
- Miscellaneous

More information

Site recode (https://seer.cancer.gov/siterecode)

Schema ID

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from NAACCR *Schema ID*
Source item number: NAACCR 3800

Schema ID links site-specific data items (SSDIs) with the appropriate primary site/histology. The values for this data item are derived based on primary site, histology, and schema discriminator fields (when required).

Considerations for use

- Schema ID is available for cases diagnosed in 2018 or later, and for specific primary site and histology groupings.
- When analyzing SSDIs (such as *Merged estrogen receptor, Merged progesterone receptor,* and *Merged HER2 summary* variables), consider using this data item in the *Selection tab* to restrict the cases to the appropriate primary sites and histology. For cases diagnosed prior to 2018, use the primary site and histology combination defined by Schema ID to restrict the cases for a comparable analysis.

More information

- Site- Specific Data Item (SSDI) Manual
- SSDI Manual Appendix A: Listing of Schema ID Information

Values

- Cervical Lymph Nodes and Unknown Primary
- Lip
- Tongue Anterior
- Gum
- Floor of Mouth
- Palate Hard
- Buccal Mucosa
- Mouth Other
- Major Salivary Glands
- Nasopharynx
- Oropharynx HPV-Mediated (p16+)
- Oropharynx (p16-)
- Hypopharynx
- Pharynx Other
- Middle Ear
- Maxillary Sinus
- Nasal Cavity and Ethmoid Sinus
- Sinus Other
- Larynx Other
- Larynx Supraglottic
- Larynx Subglottic
- Larynx Glottic
- Melanoma Head and Neck
- Cutaneous Carcinoma of Head and Neck
- Esophagus (including GE junction) Squamous
- Esophagus (including GE junction) (excluding Squamous)
- Stomach
- Small Intestine
- Appendix
- Colon and Rectum
- Anus
- Liver
- Bile Ducts Intrahepatic
- Gallbladder
- Cystic Duct
- Bile Ducts Perihilar
- Bile Duct Distal
- Ampulla of Vater
- Biliary Other
- Pancreas
- Digestive Other
- NET Stomach
- NET Duodenum
- NET Ampulla of Vater
- NET Jejunum and Ileum
- NET Appendix
- NET Colon and Rectum
- NET Pancreas
- Thymus
- Trachea
- Lung
- Retroperitoneum
- Soft Tissue Other
- Kaposi Sarcoma
- Soft Tissue Other
- Merkel Cell Skin
- Melanoma Skin
- Skin Other
- Breast
- Vulva
- Vagina
- Cervix
- Cervix (9th: 2021+)
- Cervix Sarcoma
- Corpus Carcinoma and Carcinosarcoma
- Corpus Sarcoma
- Corpus Adenosarcoma
- Ovary
- Primary Peritoneal Carcinoma
- Fallopian Tube
- Adnexa Uterine Other
- Genital Female Other
- Placenta
- Penis
- Prostate
- Testis
- Genital Male Other
- Kidney Parenchyma
- Kidney Renal Pelvis
- Bladder
- Urethra
- Urethra-Prostatic
- Urinary Other
- Skin Eyelid
- Conjunctiva
- Melanoma Conjunctiva
- Melanoma Iris
- Melanoma Choroid and Ciliary Body
- Retinoblastoma
- Lacrimal Gland
- Lacrimal Sac
- Orbital Sarcoma
- Lymphoma Ocular Adnexa
- Eye Other
- Brain
- CNS Other
- Intracranial Gland
- Thyroid
- Thyroid Medullary
- Parathyroid
- Adrenal Gland
- NET Adrenal Gland

- Pleural Mesothelioma
- Respiratory Other
- Bone Appendicular Skeleton
- Bone Spine
- Bone Pelvis
- Soft Tissue Head and Neck
- Soft Tissue Trunk and Extremities
- Soft Tissues Abdomen and Thoracic (excluding Heart, Mediastinum, Pleura)
- Heart, Mediastinum and Pleura
- GIST
- Endocrine Other
- Lymphoma
- Lymphoma-CLL/SLL
- Mycosis Fungoides
- Plasma Cell Myeloma
- Plasma Cell Disorders
- HemeRetic
- Primary Cutaneous Lymphoma (excluding MF and SS)
- Ill-Defined Other
- Blank(s)

ICCC recode 3rd edition ICD-O-3/IARC 2017

Source of standard: NCI's Surveillance, Epidemiology, and End Results (SEER) Program
Source item name: Derived from NAACCR *Primary site, Histologic code ICD-O-3,* and *Behavior code ICD-O-3*
Source item numbers: NAACCR 400 *(Primary site),* 522 *(Histologic code ICD-O-3),* and 523 *(Behavior code ICD-O-3)*

This variable indicates the classification of childhood cancer, which is based on tumor morphology and primary site. It emphasizes morphology rather than primary site, as is done for adults.

Considerations for use

NCI's SEER Program defined this variable.

Values

- I Leukemias, myeloproliferative & myelodysplastic diseases
    - I(a) Lymphoid leukemias
    - I(b) Acute myeloid leukemias
    - I(c) Chronic myeloproliferative diseases
    - I(d) Myelodysplastic syndrome and other myeloproliferative
    - I(e) Unspecified and other specified leukemias
- II Lymphomas and reticuloendothelial neoplasms
    - II(a) Hodgkin lymphomas
    - II(b) Non-Hodgkin lymphomas (except Burkitt lymphoma)
    - II(c) Burkitt lymphoma
    - II(d) Miscellaneous lymphoreticular neoplasms
    - II(e) Unspecified lymphomas
- III CNS and misc intracranial and intraspinal neoplasms
    - III(a) Ependymomas and choroid plexus tumor
    - III(b) Astrocytomas
    - III(c) Intracranial and intraspinal embryonal tumors
    - III(d) Other gliomas
    - III(e) Other specified intracranial/intraspinal neoplasms
    - III(f) Unspecified intracranial and intraspinal neoplasms
- IV Neuroblastoma and other peripheral nervous cell tumors
    - IV(a) Neuroblastoma and ganglioneuroblastoma
    - IV(b) Other peripheral nervous cell tumors
- V Retinoblastoma
- VI Renal tumors
    - VI(a) Nephroblastoma and other nonepithelial renal tumors
    - VI(b) Renal carcinomas
    - VI(c) Unspecified malignant renal tumors

- VII Hepatic tumors
    - VII(a) Hepatoblastoma
    - VII(b) Hepatic carcinomas
    - VII(c) Unspecified malignant hepatic tumors
- VIII Malignant bone tumors
    - VIII(a) Osteosarcomas
    - VIII(b) Chondrosarcomas
    - VIII(c) Ewing tumor and related sarcomas of bone
    - VIII(d) Other specified malignant bone tumors
    - VIII(e) Unspecified malignant bone tumors
- IX Soft tissue and other extraosseous sarcomas
    - IX(a) Rhabdomyosarcomas
    - IX(b) Fibrosarcomas, peripheral nerve & other fibrous
    - IX(c) Kaposi sarcoma
    - IX(d) Other specified soft tissue sarcomas
    - IX(e) Unspecified soft tissue sarcomas
- X Germ cell & trophoblastic tumors & neoplasms of gonads
    - X(a) Intracranial & intraspinal germ cell tumors
    - X(b) Extracranial & extragonadal germ cell tumors
    - X(c) Malignant gonadal germ cell tumors
    - X(d) Gonadal carcinomas
    - X(e) Other and unspecified malignant gonadal tumors
- XI Other malignant epithelial neoplasms and melanomas
    - XI(a) Adrenocortical carcinomas
    - XI(b) Thyroid carcinomas
    - XI(c) Nasopharyngeal carcinomas
    - XI(d) Malignant melanomas
    - XI(e) Skin carcinomas
    - XI(f) Other and unspecified carcinomas
- XII Other and unspecified malignant neoplasms
    - XII(a) Other specified malignant tumors
    - XII(b) Other unspecified malignant tumors
- Not classified by ICCC or *in situ*

More information

- Main Classification Table from the ICCC-3 based on ICD-O-3 (https://seer.cancer.gov/iccc/iccc3.html)
- International Classification of Childhood Cancer (ICCC) (https://seer.cancer.gov/iccc/)
- Steliarova-Foucher E, Colombet M, Ries LAG, Rous B, Stiller CA. Classification of tumours. In: Steliarova-Foucher E, Colombet M, Ries LAG, et al. *International Incidence of Childhood Cancer, Volume III.* Lyon: International Agency for Research on Cancer, In press.

ICCC recode extended 3rd edition ICD-O-3/IARC 2017

Source of standard: NCI's Surveillance, Epidemiology, and End Results Program
Source item name: Derived from NAACCR *Primary site, Histologic code ICD-O-3,* and *Behavior code ICD-O-3*
Source item numbers: NAACCR 400 *(Primary site),* 522 *(Histologic code ICD-O-3),* and 523 *(Behavior code ICD-O-3)*

This variable indicates the classification of childhood cancer, which is based on tumor morphology and primary site. It emphasizes morphology rather than primary site, as is done for adults. This variable contains extended classification codes of childhood cancer, based on definitions presented in *International Classification of Childhood Cancer, Third Edition* based on ICD-O-3/IARC 2017.

Values

- I Leukemias
  - Ia Lymphoid leukemias
    - Ia1 Precursor cell leukemias
    - Ia2 Mature B-cell leukemias
    - Ia3 Mature T-cell and NK cell leukemias
    - Ia4 Lymphoid leukemia
  - Ib Acute myeloid leukemias
  - Ic Chronic myeloproliferative diseases
  - Id Myelodysplastic syndrome and other myeloproliferative diseases
  - Ie Unspecified and other specified leukemias
- II Lymphomas and reticuloendothelial neoplasms
  - IIa Hodgkin lymphomas
  - IIb Non-Hodgkin lymphomas (except Burkitt lymphoma)
    - IIb1 Precursor cell lymphomas
    - IIb2 Mature B-cell lymphomas (except Burkitt lymphoma)
    - IIb3 Mature T-cell and NK-cell lymphomas
    - IIb4 Non-Hodgkin lymphomas
  - IIc Burkitt lymphoma
  - IId Miscellaneous lymphoreticular neoplasms
  - IIe Unspecified lymphomas
- III CNS and miscellaneous intracranial and intraspinal neoplasms
  - IIIa Ependymomas and choroid plexus tumor
    - IIIa1 Ependymomas
    - IIIa2 Choroid plexus tumor
  - IIIb Astrocytomas
  - IIIc Intracranial and intraspinal embryonal tumors
    - IIIc1 Medulloblastomas
    - IIIc2 Primitive neuroectodermal tumor (PNET)
    - IIIc3 Medulloepithelioma
    - IIIc4 Atypical teratoid/rhabdoid tumor
  - IIId Other gliomas
    - IIId1 Oligodendrogliomas
    - IIId2 Mixed and unspecified gliomas
    - IIId3 Neuroepithelial glial tumors of uncertain origin
  - IIIe Other specified intracranial and intraspinal neoplasms
    - IIIe1 Pituitary adenomas and carcinomas
    - IIIe2 Tumours of the sellar region (craniopharyngiomas)
    - IIIe3 Pineal parenchymal tumors
    - IIIe4 Neuronal and mixed neuronal-glial tumors
    - IIIe5 Meningiomas
  - IIIf Unspecified intracranial and intraspinal neoplasms
- IV Neuroblastoma and other peripheral nervous cell tumors
  - IVa Neuroblastoma and ganglioneuroblastoma
  - IVb Other peripheral nervous cell tumors
- V Retinoblastoma
- VI Renal tumors
  - VIa Nephroblastoma and other non-epithelial renal tumors
    - VIa1 Nephroblastoma
    - VIa2 Rhabdoid renal tumor
    - VIa3 Kidney sarcomas
  - VIb Renal carcinomas

    - VIc Unspecified malignant renal tumors
- VII Hepatic tumors
    - VIIa Hepatoblastoma and mesenchymal tumors of liver
        - VIIa1 Hepatoblastoma
        - VIIa2 Rhabdoid hepatic tumor
        - VIIa3 Embryonal sarcoma of liver
    - VIIb Hepatic carcinomas
    - VIIc Unspecified malignant hepatic tumors
- VIII Malignant bone tumors
    - VIIIa Osteosarcomas
    - VIIIb Chondrosarcomas
    - VIIIc Ewing tumor and related sarcomas of bone
        - VIIIc1 Ewing tumor and Askin tumor of bone
        - VIIIc2 Peripheral neuroectodermal tumor (pPNET) of bone
    - VIIId Other specified malignant bone tumors
        - VIIId1 Malignant fibrous neoplasms of bone
        - VIIId2 Malignant chordomas
        - VIIId3 Odontogenic malignant tumors
        - VIIId4 Miscellaneous malignant bone tumors
    - VIIIe Unspecified malignant bone tumors
- IX Soft tissue and other extraosseous sarcomas
    - IXa Rhabdomyosarcomas
    - IXb Fibrosarcomas
        - IXb1 Fibroblastic and myofibroblastic tumors
        - IXb2 Nerve sheath tumors
        - IXb3 Other fibromatous neoplasms
    - IXc Kaposi sarcoma
    - IXd Other specified soft tissue sarcomas
        - IXd1 Ewing tumor and Askin tumor of soft tissue
        - IXd2 Peripheral neuroectodermal tumor (pPNET) of soft tissue
        - IXd3 Extrarenal extrahepatic rhabdoid tumor
        - IXd4 Liposarcomas
        - IXd5 Fibrohistiocytic tumors
        - IXd6 Leiomyosarcomas
        - IXd7 Synovial sarcomas
        - IXd8 Blood vessel tumors
        - IXd9 Osseous and chondromatous neoplasms of soft tissue
        - IXd10 Alveolar soft parts sarcoma
        - IXd11 Miscellaneous soft tissue sarcomas
    - IXe Unspecified soft tissue sarcomas
- X Germ cell tumors
    - Xa Intracranial and intraspinal germ cell tumors
        - Xa1 Intracranial and intraspinal germinomas
        - Xa2 Intracranial and intraspinal teratomas
        - Xa3 Intracranial and intraspinal embryonal carcinomas
        - Xa4 Intracranial and intraspinal yolk sac tumor
        - Xa5 Intracranial and intraspinal choriocarcinoma
        - Xa6 Intracranial and intraspinal tumors of mixed forms
    - Xb Malignant extracranial and extragonadal germ cell tumors
        - Xb1 Malignant germinomas of extracranial and extragonadal sites
        - Xb2 Malignant teratomas of extracranial and extragonadal sites
        - Xb3 Embryonal carcinomas of extracranial and extragonadal sites

        - Xb4 Yolk sac tumor of extracranial and extragonadal sites
        - Xb5 Choriocarcinomas of extracranial and extragonadal sites
        - Xb6 Oth/unspec malig mixed germ cell tumors of extracranial/extragonadal
    - Xc Malignant gonadal germ cell tumors
        - Xc1 Malignant gonadal germinomas
        - Xc2 Malignant gonadal teratomas
        - Xc3 Gonadal embryonal carcinomas
        - Xc4 Gonadal yolk sac tumor
        - Xc5 Gonadal choriocarcinoma
        - Xc6 Malignant gonadal tumors of mixed forms
        - Xc7 Malignant gonadal gonadoblastoma
    - Xd Gonadal carcinomas
    - Xe Other and unspecified malignant gonadal tumors
- XI Other malignant epithelial neoplasms and malignant melanomas
    - XIa Adrenocortical carcinomas
    - XIb Thyroid carcinomas
    - XIc Nasopharyngeal carcinomas
    - XId Malignant melanomas
    - XIe Skin carcinomas
    - XIf Other and unspecified carcinomas
        - XIf1 Carcinomas of salivary glands
        - XIf2 Carcinomas of colon and rectum
        - XIf3 Carcinomas of appendix
        - XIf4 Carcinomas of lung
        - XIf5 Carcinomas of thymus
        - XIf6 Carcinomas of breast
        - XIf7 Carcinomas of cervix uteri
        - XIf8 Carcinomas of bladder
        - XIf9 Carcinomas of eye
        - XIf10 Carcinomas of other specified sites
        - XIf11 Carcinomas of unspecified site
- XII Other and unspecified malignant neoplasms
    - XIIa Other specified malignant tumors
        - XIIa1 Malignant gastrointestinal stromal tumor
        - XIIa2 Pancreatoblastoma
        - XIIa3 Pulmonary blastoma and pleuropulmonary blastoma
        - XIIa4 Other complex mixed and stromal neoplasms
        - XIIa5 Mesothelioma
        - XIIa6 Other specified malignant tumors
    - XIIb Other unspecified malignant tumors
- Not classified by SEER or *in situ*

More information

- Main Classification Table from the ICCC-3 based on ICD-O-3 (https://seer.cancer.gov/iccc/iccc3_ext.html)
- International Classification of Childhood Cancer (ICCC) (https://seer.cancer.gov/iccc/)
- Steliarova-Foucher E, Colombet M, Ries LAG, Rous B, Stiller CA. Classification of tumours. In: Steliarova-Foucher E, Colombet M, Ries LAG, et al. *International Incidence of Childhood Cancer, Volume III.* Lyon: International Agency for Research on Cancer, In press.

AYA site recode 2020 revision

Source of standard: NCI's Surveillance, Epidemiology, and End Results Program

Source item name: Derived from NAACCR *Primary site, Histologic code ICD-O-3,* and *Behavior code ICD-O-3*
Source item numbers: NAACCR 400 *(Primary site),* 522 *(Histologic code ICD-O-3),* and 523 *(Behavior code ICD-O-3)*

This variable was developed to define the major cancer sites that affect adolescents and young adults (AYAs) between 15 and 39 years of age.

Considerations for use

The SEER Program defined this recode variable based on the classification scheme proposed by RD Barr and colleagues. Refer to the AYA Site Recode at https://seer.cancer.gov/ayarecode/ for the full list of 318 groups and more information.

Values

- Leukemias and related disorders
- Lymphomas
- CNS and other intracranial and intraspinal neoplasms
- Sarcomas
- Blood and lymphatic vessel tumors
- Nerve sheath tumors
- Gonadal and related tumors
- Melanoma – malignant
- Carcinomas
- Miscellaneous specified neoplasms
- Unclassified and Non-malignant

More information

- AYA Site Recode (https://seer.cancer.gov/ayarecode)
- Barr RD, Holowaty EJ, Birch JM. Classification schemes for tumors diagnosed in adolescents and young adults. *Cancer.* 2006;106(7):1425–1430.

Lymphoid neoplasm recode 2021 revision

Source of standard: NCI's Surveillance, Epidemiology, and End Results (SEER) Program
Source item name: Derived from NAACCR *Primary site, Histologic code ICD-O-3,* and *Behavior code ICD-O-3*
Source item numbers: NAACCR 400 *(Primary site),* 522 *(Histologic code ICD-O-3),* and 523 *(Behavior code ICD-O-3)*

This variable was based on ICD-O-3, updated for hematopoietic codes based on *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (2008).* It was designed to facilitate epidemiologic studies of lymphoma subtypes.

Considerations for use

- The SEER Program defined this recode variable. It was adapted from a proposed nested classification of lymphoid neoplasms in Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph) (2007) (https://pmc.ncbi.nlm.nih.gov/articles/PMC1924473/).

- Refer to SEER's Lymphoid neoplasm subtype recode 2020 documentation at https://seer.cancer.gov/lymphomarecode/lymphoma-2020.html for the full list of categories and more information.

Values

- Lymphoid Neoplasm
- 1 Hodgkin Lymphoma
    - 1(a) Classical Hodgkin lymphoma
        - 1(a)1 Lymphocyte-rich/mixed cell/lymphocyte depleted

- 1(a)1.1 Lymphocyte-rich
- 1(a)1.2 Mixed cellularity
- 1(a)1.3 Lymphocyte-depleted

- 1(a)2 Nodular sclerosis
- 1(a)3 Classical Hodgkin lymphoma, NOS

- 1(b) Nodular lymphocyte predominant Hodgkin lymphoma

- 2 Non-Hodgkin Lymphoma
  - 2(a) Non-Hodgkin lymphoma, B-cell
    - 2(a)1 Precursor Non-Hodgkin lymphoma, B-cell
    - 2(a)2 Mature Non-Hodgkin lymphoma, B-cell
      - 2(a)2.1 Chronic/Sm/Prolymphocytic/Mantle B-cell NHL
        - 2(a)2.1.1 Chronic/Small lymphocytic leuk/lymph
        - 2(a)2.1.2 Prolymphocytic leukemia, B-cell
        - 2(a)2.1.3 Mantle-cell lymphoma
      - 2(a)2.2 Lymphoplasmacytic lymphoma/Waldenstrom
        - 2(a)2.2.1 Lymphoplasmacytic lymphoma
        - 2(a)2.2.2 Waldenstrom macroglubulinemia
      - 2(a)2.3 Diffuse large B-cell lymphoma (DLBCL)
        - 2(a)2.3.1 DLBCL, NOS
        - 2(a)2.3.2 Intravascular large B-cell lymphoma
        - 2(a)2.3.3 Primary effusion lymphoma
        - 2(a)2.3.4 Mediastinal large B-cell lymphoma
      - 2(a)2.4 Burkitt lymphoma/leukemia
      - 2(a)2.5 Marginal-zone lymphoma (MZL)
        - 2(a)2.5.1 Splenic MZL
        - 2(a)2.5.2 Extranodal MZL, MALT type
        - 2(a)2.5.3 Nodal MZL
      - 2(a)2.6 Follicular lymphoma
      - 2(a)2.7 Hairy-cell leukemia
      - 2(a)2.8 Plasma cell neoplasms
        - 2(a)2.8.1 Plasmacytoma
        - 2(a)2.8.2 Multiple myeloma/plasma-cell leuk
      - 2(a)2.9 Heavy chain disease
    - 2(a)3 Non-Hodgkin lymphoma, B-cell, NOS
  - 2(b) Non-Hodgkin lymphoma, T-cell
    - 2(b)1 Precursor Non-Hodgkin lymphoma, T-cell
    - 2(b)2 Mature Non-Hodgkin lymphoma, T-cell
      - 2(b)2.1 Mycosis fungoides/Sezary syndrome
        - 2(b)2.1.1 Mycosis fungoides
        - 2(b)2.1.2 Sezary syndrome
    - 2(b)2.2 Peripheral T-cell lymphoma
        - 2(b)2.2.1 Peripheral T-cell lymphoma, NOS
        - 2(b)2.2.2 Angioimmunoblastic T-cell lymphoma
        - 2(b)2.2.3 Subcutan panniculitis-like T-cell lymph
        - 2(b)2.2.4 Anaplastic lar cell lymph, T-/Null-cell
        - 2(b)2.2.5 Hepatosplenic T-cell lymphoma
        - 2(b)2.2.6 Enteropathy-type T-cell lymphoma
        - 2(b)2.2.7 Cutaneous T-cell lymphoma, NOS
        - 2(b)2.2.8 Prim cutaneous anaplastic lar cell lymph
    - 2(b)2.3 Adult T-cell leukemia/lymphoma
    - 2(b)2.4 NK/T-cell lymph, nasal-type/aggres NK leuk
    - 2(b)2.5 T-cell large granular lymphocytic leukemia

        - 2(b)2.6 Prolymphocytic leukemia, T-cell
    - 2(c) Non-Hodgkin lymphoma, unknown lineage
- 3 Composite Hodgkin lymphoma and NHL
- 4 Lymphoid neoplasm, NOS
- Unclassified

More information

Lymphoid neoplasm subtype recode 2020 (https://seer.cancer.gov/lymphomarecode/lymphoma-2020.html)

Variable Definitions: Stage – local, regional, distant (LRD) [summary and historic]

Merged summary stage

Source of standard: CDC's National Program of Cancer Registries
Source item name: Combined from *Derived SS2000* and *SEER Summary Stage 2000*
Source item numbers: Derived from NAACCR 3020 *(Derived SS2000),* 759 *(SEER Summary Stage 2000),* and 764 *(Summary Stage 2018)*

This is a merged stage variable created using the variables *SEER Summary Stage 2000, Derived SS2000,* and *Summary Stage 2018*. This stage variable can be used for diagnosis years 2001–2022.

Considerations for use

This variable is not available for testis cases. The coding logic for this merged variable is as follows:

For NPCR-registries

- If a case was diagnosed in 2001, 2002, 2003, 2016 or 2017, stage at diagnosis is recorded using the *SEER Summary Stage 2000* variable value.
- If a case was diagnosed in or between 2004 and 2015, stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value. If the *Derived SEER Summary Stage 2000* variable is blank or unstaged, and the *SEER Summary Stage 2000* variable has a valid value, that value is used to populate the merged variable.
- If a case was diagnosed in 2018 or later, stage at diagnosis is recorded using the *Summary Stage 2018* variable value.

For SEER-only registries (Connecticut, Hawaii, Iowa, and New Mexico)

- If a case was diagnosed in 2001, 2002, or 2003, stage at diagnosis is recorded using the *SEER Summary Stage 2000* variable value.
- If a case was diagnosed in or between 2004 and 2017, stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value.
- If a case was diagnosed in 2018 or later, stage at diagnosis is recorded using the *Derived Summary Stage 2018* variable value.

Values

- *In situ*
- Localized only
- Regional, direct extension only
- Regional, regional lymph nodes only
- Regional, direct extension and regional lymph nodes
- Regional, NOS
- Distant site(s)/node(s) involved
- Benign/borderline
- Not applicable
- Unknown/unstaged/unspecified/DCO

More information

SEER Summary Stage 2000, Derived SS2000, and *Summary Stage 2018* variables

Rx summary – surgery primary site

Source of standard: NCI's Surveillance, Epidemiology, and End Results (SEER) Program and Commission on Cancer
Source item name: *RX Summ—Surg Prim Site*
Source item number: NAACCR 1290

This variable records site-specific codes for the type of surgery to the primary site performed as part of the first course of treatment. This includes treatment given at all facilities as part of the first course of treatment.

Considerations for use

Data for this variable are available for **female breast cancer cases** starting with **diagnosis year 2010.**

Values

- 00–99

More information

- *SEER Program Coding and Staging Manual 2016, Appendix C: Site Specific Coding Modules, Breast Surgery Codes* (https://seer.cancer.gov/archive/manuals/2016/AppendixC/Surgery_Codes_Breast_2016.pdf)
- Facility Oncology Registry Data Standards (FORDS)
- SEER Program Coding and Staging Manual 2018 (https://seer.cancer.gov/tools/codingmanuals)

Variable Definitions: Extent of Disease

CS site-specific factor 1

Source of standard: American Joint Committee on Cancer
Source item name: *CS Site-Specific Factor 1*
Source item number: NAACCR 2880

The information recorded in this variable differs for each anatomic site. There are site-specific codes and coding structures for each anatomic site. In the U.S. Cancer Statistics Public Use Database, this variable records the World Health Organization (WHO) Grade Classification for brain and other nervous system sites.

Considerations for use

Data for this variable are available for brain and other nervous system cases diagnosed from 2011 through 2017.

Collection of Collaborative Stage (CS) Site-Specific Factor data items stopped in 2017. The information is now collected in the *Grade Pathological* and *Grade Clinical* variables for 2018 or later cases.

For the site-specific codes, please refer to the Collaborative Stage Data Collection System, brain cancer: World Health Organization (WHO) Grade Classification.

Values

- 010–999
- Blank(s)

Merged estrogen receptor

Source of standard: CDC's National Program of Cancer Registries
Source item name: Combined from *CS Site-Specific Factor 1 (breast)* and *Estrogen Receptor Summary*
Source item number: Derived from NAACCR 2880 *(CS Site-Specific Factor 1)* and 3827 *(Estrogen Receptor Summary from Site-Specific Data Item)*

This is a merged variable created using the variables *CS Site-Specific Factor 1 (breast)* and *Estrogen Receptor Summary* and is the summary of results of the estrogen receptor (ER) assay.

Considerations for use

- Data for this variable are available for female breast cancer cases diagnosed in 2004 or later.
- When using this data item, restrict the query to the appropriate cases using the *Schema ID* variable for cases diagnosed in 2018 onwards. For cases diagnosed prior to 2018, use the primary site and histology combination defined by *Schema ID* to restrict the cases for a comparable analysis.

Values

- ER negative
- ER positive
- Test ordered, results not in chart
- Not documented, indetermine, unknown
- Blank(s)

Merged progesterone receptor

Source of standard: CDC's National Program of Cancer Registries
Source item name: Combined from *CS Site-Specific Factor 2 (breast)* and *Progesterone Receptor Summary*
Source item number: Derived from NAACCR 2890 *(CS Site-Specific Factor 2)* and 3915 *(Progesterone Receptor Summary from Site-Specific Data Item)*

This is a merged variable created using the variables *CS Site-Specific Factor 2 (breast)* and *Progesterone Receptor Summary* and is the summary of results of the progesterone receptor (PR) assay.

Considerations for use

- Data for this variable are available for female breast cancer cases diagnosed in 2010 or later.
- When using this data item, restrict the query to the appropriate cases using the *Schema ID* variable for cases diagnosed in 2018 onwards. For cases diagnosed prior to 2018, use the primary site and histology combination defined by *Schema ID* to restrict the cases for a comparable analysis.

Values

- PR negative
- PR positive
- Test ordered, results not in chart
- Not documented, indetermine, unknown
- Blank(s)

Merged HER2 summary

Restricted to female breast and diagnosis years 2010 or later

Source of standard: CDC's National Program of Cancer Registries
Source item name: Combined from *CS Site-Specific Factor 15 (breast)* and *HER2 Overall Summary*
Source item number: Derived from NAACCR 2869 *(CS Site-Specific Factor 15)* and 3855 *(HER2 Overall Summary from Site-Specific Data Item for breast)*

This is a merged variable created using the variables *CS Site-Specific Factor 15 (breast)* and *HER2 Overall Summary* and is the summary of results from HER2 testing.

Considerations for use

- Data for this variable are available for female breast cancer cases diagnosed in 2011 or later.
- When using this data item, restrict the query to the appropriate cases using the *Schema ID* variable for cases diagnosed in 2018 onwards. For cases diagnosed prior to 2018, use the primary site and histology combination defined by *Schema ID* to restrict the cases for a comparable analysis.

Values

- HER2 negative, equivocal
- HER2 positive
- Test ordered, results not in chart
- Not documented, indetermine, unknown
- Blank(s)

Variable Definitions: Multiple Primary Fields

Sequence number – central

Source of standard: North American Association of Central Cancer Registries
Source item name: *Sequence Number – Central*
Source item number: 380

This variable indicates the sequence of all reportable neoplasms over the patient's lifetime.

Considerations for use

- The sequence number may change over the patient's lifetime. If the patient was diagnosed with a single reportable neoplasm, and later diagnosed with a second reportable neoplasm, the sequence code for the first neoplasm changes from 00 to 01. A central registry may find that a patient with one or more known neoplasms had an earlier reportable neoplasm that had been unknown to the registry. Typically, a re-evaluation of all related sequence numbers is required whenever an additional neoplasm is identified.

- Standards define which neoplasms are reportable. It is assumed that these standards are the minimum definition of reportability. Individual central cancer registries may define additional neoplasms as reportable. Variability of assigning sequence numbers over time may exist for different registries, which may impact the coding of this variable.

- Because the time period of *Sequence Number* is a person's lifetime, reportable neoplasms not included in the central registry (those that occur outside the registry catchment area or before the reference date) also are allotted a sequence number. For example, a registry may contain a single record for a patient with a sequence number of 02 because the first reportable neoplasm preceded the central registry's reference date.

- If two or more reportable neoplasms are diagnosed at the same time, the lowest sequence number is assigned to the diagnosis with the worst prognosis. If no difference in prognosis is evident, the decision is arbitrary.

- Reportable non-malignant tumors diagnosed on or after January 1, 2004 are represented by sequence numbers labeled as "…state registry-defined neoplasm". Timing rules for sequencing these neoplasms are the same as timing rules for sequencing required *in situ* or invasive neoplasms.

- The 2007 Multiple Primary and Histology Coding Rules (https://seer.cancer.gov/tools/mphrules/mphrules_instructions.pdf) may also affect the sequence number.

Values

- One primary only
- 1st of 2 or more primaries
- 2nd of 2 or more primaries
- …
- 59th of 59 or more primaries
- Only one state registry-defined neoplasm
- 1st of 2 or more state registry-defined neoplasms
- 2nd of 2 or more state registry-defined neoplasms
- …
- 20th of 20 or more state registry-defined neoplasms
- Unknown sequence number - state registry-defined neoplasms
- Carcinoma *in situ* of the Cervix diagnosed 1/1/1996 or later
- Unknown sequence number - federally required *in situ* or malignant tumors

More information

- SEER Program Coding and Staging Manual (https://seer.cancer.gov/tools/codingmanuals)
- SEER Solid Tumor Rules Manual, 2024 (https://seer.cancer.gov/tools/solidtumor/)

Variable Definitions: Dates

Year of birth

Source of standard: NCI's Surveillance, Epidemiology, and End Results (SEER) Program and Commission on Cancer
Source item name: *Date of Birth*
Source item number: 240

The patient's year of birth.

Considerations for use

- The month and day of birth are not provided for confidentiality reasons.
- If age at diagnosis and year of diagnosis are known, but year of birth is unknown, the year of birth is calculated and so coded. Only the year is entered. Per the NAACCR Data Dictionary, registrars are instructed to estimate a date of birth rather than leave the birth date unknown.
- **This variable includes only count data.** Rates cannot be calculated using this variable as no population data are associated with it.

Values

- 1890
- 1891
- …
- 2022
- Blank(s)

Month of diagnosis

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from *Date of Diagnosis*
Source item number: 390

This variable is derived from *Date of initial diagnosis*, which is the date of initial diagnosis by a recognized medical practitioner for the cancer being reported, whether clinically or microscopically confirmed.

Considerations for use

- The day of diagnosis is not provided as an additional confidentiality measure.
- **This variable includes only count data.** Rates cannot be calculated using this variable as no population data are associated with it.

Values

- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December
- Blank(s)
- Invalid value(s)

Rural-urban continuum 2013, grouped

Source of standard: North American Association of Central Cancer Registries
Source item name: Derived from *RuralUrban Continuum 2013*
Source item number: 3312

The U.S. Department of Agriculture Economic Research Service's 2013 Rural-Urban Continuum Codes form a classification scheme that distinguishes metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area.

This variable in this database groups the 2013 Rural-Urban Continuum Codes (also referred to as the Beale Codes) into 3 categories: metropolitan counties (rural-urban continuum codes 1–3), nonmetropolitan counties (rural-urban continuum codes 4–9), and unavailable (blank or unknown).

Categorizing counties by population size helps researchers investigate geographic correlates of the burden of cancer in the area of interest.

Considerations for use

These codes are derived electronically by the central cancer registry using patients' county at diagnosis.

More information

U.S. Department of Agriculture Economic Research Service's 2013 Rural-Urban Continuum Codes (www.ers.usda.gov/data-products/rural-urban-continuum-codes)

Values

- Metropolitan
- Non-Metropolitan
- Unavailable

Variable Definitions: Merged System-Supplied

Alcohol-related cancers

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Primary Site, Histologic Type ICD-O-3,* and *Sex*
Source item numbers: Derived from NAACCR's 400 *(Primary Site),* 522 *(Histologic Type ICD-O-3),* and 220 *(Sex)*

This is a predefined variable created using ICD-O-3 site, histology, and sex to define alcohol-related cancers.[1,2]

Considerations for use

- Cancer registries do not routinely collect data on alcohol use, so the number of cancers associated with this risk factor cannot be determined definitively.[3,4,5]
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, known as the *attributable fraction.*[6] The number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, please see the referenced publications and Definitions of Risk Factor-Associated Cancers documentation at www.cdc.gov/united-states-cancer-statistics/public-use/definitions-risk-factor-associated-cancers.html.
- Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers except for bladder cancers (includes *in situ* and invasive cancers).

Values

- Lip, oral cavity, & pharynx
- Esophagus
- Colon & rectum
- Liver
- Larynx
- Female breast

References

1. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans: Volume 96: Alcohol Consumption and Ethyl Carbamate. Lyon, France: International Agency for Research on Cancer; 2010.
2. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans: Volume 100E: Personal Habits and Indoor Combustions: Consumption of Alcoholic Beverages. Lyon, France: International Agency for Research on Cancer; 2012.
3. Henley SJ, Singh SD, King J, Wilson RJ, O'Neil ME, Ryerson AB. Invasive cancer incidence and survival—United States, 2013. *MMWR Morb Mortal Wkly Rep.* 2017;66:69–75.
4. Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst.* 2011;103(24):1827–1839.
5. World Cancer Research Fund / American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective.* Washington, DC: AICR, 2007.
6. Levine B. What does the population attributable fraction mean? *Prev Chronic Dis.* 2007;4(1):A14.

Human papillomavirus (HPV)-related cancers

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Primary Site, Histologic Type ICD-O-3, Sex,* and *Diagnostic Confirmation*
Source item numbers: Derived from NAACCR's 400 *(Primary Site),* 522 *(Histologic Type ICD-O-3),*220 *(Sex),* and 490 *(Diagnostic Confirmation)*

This is a predefined variable created using ICD-O-3 site, histology, and sex to define human papillomavirus (HPV)-related cancers.[1,2,3,4]

Considerations for use

- Cancer registries do not routinely collect data on HPV diagnoses, so the number of cancers associated with this risk factor cannot be determined definitively.[5,6,7]
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, known as the *attributable fraction.*[8] The number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, please see the referenced publications and Definitions of Risk Factor-Associated Cancers documentation at www.cdc.gov/united-states-cancer-statistics/public-use/definitions-risk-factor-associated-cancers.html.
- Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers except for bladder cancers (includes *in situ* and invasive cancers).

Values

- Oropharyngeal squamous cell carcinoma
- Anal and rectal squamous cell carcinoma
- Vulvar squamous cell carcinoma
- Vaginal squamous cell carcinoma
- Penile squamous cell carcinoma
- Cervical carcinoma

References

1. Watson M, Saraiya M, Ahmed F, et al. Using population-based cancer registry data to assess the burden of human papillomavirus-associated cancers in the United States: overview of methods. *Cancer.* 2008;113(10 Suppl):2841–2854.
2. Saraiya M, Unger ER, Thompson TD, et al. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. *J Natl Cancer Inst.* 2015;107(6):djv086.
3. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans. Volume 90: Human Papillomaviruses. Lyon, France: International Agency for Research on Cancer; 2007.
4. Viens LJ, Henley SJ, Watson M, et al. Human papillomavirus–associated cancers—United States, 2008–2012. *MMWR Morb Mortal Wkly Rep.* 2016;65(26):661–666.
5. Henley SJ, Singh SD, King J, Wilson RJ, O'Neil ME, Ryerson AB. Invasive cancer incidence and survival—United States, 2013. *MMWR Morb Mortal Wkly Rep.* 2017;66:69–75.
6. Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst.* 2011;103(24):1827–1839.

7. World Cancer Research Fund / American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*. Washington, DC: AICR, 2007.

8. Levine B. What does the population attributable fraction mean? *Prev Chronic Dis.* 2007;4(1):A14.

Obesity-related cancers

Source of standard: CDC's National Program of Cancer Registries

Source item name: Derived from NAACCR's *Primary Site, Histologic Type ICD-O-3, Sex, Diagnostic Confirmation,* and *Age at Diagnosis*

Source item numbers: Derived from NAACCR's 400 *(Primary Site),* 522 *(Histologic Type ICD-O-3),* 220 *(Sex),* 490 *(Diagnostic Confirmation),* and 230 *(Age at diagnosis)*

This is a predefined variable created using ICD-O-3 site, histology, and sex to define obesity-related cancers.[1,2,3]

Considerations for use

- Cancer registries do not routinely collect data on obesity, so the number of cancers associated with this risk factor cannot be determined definitively.[2,4,5]
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, known as the *attributable fraction*.[6] The number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, please see the referenced publications and Definitions of Risk Factor-Associated Cancers documentation at www.cdc.gov/united-states-cancer-statistics/public-use/definitions-risk-factor-associated-cancers.html.
- Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers except for bladder cancers (includes *in situ* and invasive cancers).

Values

- Esophageal adenocarcinoma
- Gastric cardia
- Colon & rectum
- Liver
- Gallbladder
- Pancreas
- Kidney
- Meningioma
- Thyroid
- Multiple myeloma
- Post-menopausal female breast cancer
- Corpus & uterus NOS
- Ovary

References

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

2. World Cancer Research Fund / American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*. Washington, DC: AICR, 2007.

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

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

5. Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst.* 2011;103(24):1827–1839.

6. Levine B. What does the population attributable fraction mean? *Prev Chronic Dis.* 2007;4(1):A14.

Physical inactivity-related cancers

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Primary Site, Histologic Type ICD-O-3,* and *Sex*
Source item numbers: Derived from NAACCR's 400 *(Primary Site),* 522 *(Histologic Type ICD-O-3),* and 220 *(Sex)*

This is a predefined variable created using ICD-O-3 site, histology, and sex to define physical inactivity-related cancers.[1,2]

Considerations for use

- Cancer registries do not routinely collect data on physical inactivity, so the number of cancers associated with this risk factor cannot be determined definitively.[2,3,4]
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, known as the *attributable fraction.*[5] The number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, please see the referenced publications and Definitions of Risk Factor-Associated Cancers documentation at www.cdc.gov/united-states-cancer-statistics/public-use/definitions-risk-factor-associated-cancers.html.
- Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers except for bladder cancers (includes *in situ* and invasive cancers).

Values

- Colon
- Post-menopausal female breast
- Corpus and uterus NOS

References

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

2. World Cancer Research Fund / American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective.* Washington, DC: AICR, 2007.

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

4. Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst.* 2011;103(24):1827–1839.

5. Levine B. What does the population attributable fraction mean? *Prev Chronic Dis.* 2007;4(1):A14.

Tobacco-related cancers

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from NAACCR's *Primary Site, Histologic Type ICD-O-3,* and *Sex*
Source item numbers: Derived from NAACCR's 400 *(Primary Site),* 522 *(Histologic Type ICD-O-3),* and 220 *(Sex)*

This is a predefined variable created using ICD-O-3 site, histology, and sex to define tobacco-related cancers.[1]

Considerations for use

- Cancer registries do not routinely collect data on tobacco use, so the number of cancers associated with this risk factor cannot be determined definitively.[2,3,4]
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, known as the *attributable fraction.*[5] The number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, please see the referenced publications and Definitions of Risk Factor-Associated Cancers documentation at www.cdc.gov/united-states-cancer-statistics/public-use/definitions-risk-factor-associated-cancers.html.
- Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers except for bladder cancers (includes *in situ* and invasive cancers).

Values

- Lip, oral cavity, & pharynx
- Esophagus
- Stomach
- Colon & rectum
- Liver
- Pancreas
- Larynx
- Trachea, lung, & bronchus
- Cervix uteri
- Kidney & renal pelvis
- Urinary bladder
- Acute myeloid leukemia

References

1. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.
2. Henley SJ, Singh SD, King J, Wilson RJ, O'Neil ME, Ryerson AB. Invasive cancer incidence and survival—United States, 2013. *MMWR Morb Mortal Wkly Rep.* 2017;66:69–75.
3. Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst.* 2011;103(24):1827–1839.
4. World Cancer Research Fund / American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective.* Washington, DC: AICR, 2007.
5. Levine B. What does the population attributable fraction mean? *Prev Chronic Dis.* 2007;4(1):A14.

State race ethnicity suppress

Source of standard: CDC's National Program of Cancer Registries
Source item name: Derived from *Addr at DX – state* and state-level race or ethnicity reporting restrictions
Source item number: Derived from NAACCR's 80

This variable was created specifically for this database. It provides the selection of states that are eligible to be included in a state-level analysis of race and ethnicity combined.

Considerations for use

- States have the option to suppress race-specific and Hispanic ethnicity-specific data. While these states can be included in an aggregated analysis, the affected state's information cannot be reported at the state level.
- Use this variable when conducting state-level analyses of race and ethnicity combinations. If you are conducting a state-level analysis of race or ethnicity only, make restrictions manually in the SEER*Stat Selection tab.
- The following states have state-level race or ethnicity data presentation restrictions:
    - Data for Hispanic and non-Hispanic American Indian and Alaska Native people cannot be displayed for Illinois, Kansas, New Jersey, and New York.
    - Data for Hispanic Asian and Pacific Islander and Hispanic Black people cannot be displayed for Kansas.
- For more information, please refer to the *Race recode (W, B, AIAN, API), Origin recode NHIA (Hispanic, Non-Hisp),* and *Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)* variable descriptions.

Values

- All races
- White
- White Non-Hispanic
- White Hispanic
- Black
- Black Non-Hispanic
- Black Hispanic

Definitions of Risk Factor-Associated Cancers

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

Cancer registries do not routinely collect risk factor information. So estimates for risk factor-associated cancers are based only on cancer type.

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

The following tables list the International Classification of Disease for Oncology (ICD-O)-3 site codes, histology codes, and additional restrictions used to define and code the predefined SEER*Stat variables. Further restrictions may be made depending on the analysis. Refer to the SEER*Stat tutorial on merged variables at https://seer.cancer.gov/seerstat/tutorials/rate3/webprint/ for information on editing a merged variable.

Alcohol-associated cancers[2,7]

Cancer	ICD-O-3 site codes	ICD-O-3 histology codes	Additional restrictions
Lip, oral cavity, pharynx	C00.0–14.8	8000–9049, 9056–9139, 9141–9589	
Esophagus	C15.0–15.9	8000–9049, 9056–9139, 9141–9589	
Colon and rectum	C18.0–20.9, C26.0	8000–9049, 9056–9139, 9141–9589	
Liver	C22.0	8000–9049, 9056–9139, 9141–9589	
Larynx	C32.0–32.9	8000–9049, 9056–9139, 9141–9589	
Female breast	C50.0–50.9	8000–9049, 9056–9139, 9141–9589	Restrict to females

Human papillomavirus-associated cancers[3,8,9,10,11]

Cancer	ICD-O-3 site codes	ICD-O-3 histology codes	Additional restrictions
Oropharyngeal squamous cell carcinoma	C01.9, 02.4, 02.8, 05.1–05.2, 09.0–09.1, 09.8–09.9, 10.0–10.4, 10.8–10.9, 14.0, 14.2, 14.8	8050–8086, 8120–8131	Restrict to microscopically confirmed
Anal and rectal squamous cell carcinoma	C21.0–21.8, 20.9	8050–8086, 8120–8131	Restrict to microscopically confirmed
Vulvar squamous cell carcinoma	C51.0–51.9	8050–8086, 8120–8131	Restrict to females and restrict to microscopically confirmed
Vaginal squamous cell carcinoma	C52.9	8050–8086, 8120–8131	Restrict to females and restrict to microscopically confirmed
Cervical carcinoma	C53.0–53.9	8010–8671, 8940–8941	Restrict to females and restrict to microscopically confirmed
Penile squamous cell carcinoma	C60.0–60.9	8050–8086, 8120–8131	Restrict to males and restrict to microscopically confirmed

Obesity-associated cancers[4,5,12]

Cancer	ICD-O-3 site codes	ICD-O-3 histology codes	Additional restrictions
Esophageal adenocarcinoma	C15.0–15.9	8140–8575	Restrict to microscopically confirmed
Gastric cardia	C16.0	8000-9049, 9056–9139, 9141–9589	
Colon and rectum	C18.0–20.9, C26.0	8000–9049, 9056–9139, 9141–9589	
Liver	C22.0	8000–9049, 9056–9139, 9141–9589	
Gallbladder	C23.9	8000–9049, 9056–9139, 9141–9589	
Pancreas	C25.0–25.9	8000–9049, 9056–9139, 9141–9589	
Multiple myeloma	C42.1	9732	
Postmenopausal female breast	C50.0–50.9	8000–9049, 9056–9139, 9141–9589	Restrict to females and restrict to age ≥50 years
Corpus and uterus, NOS (not otherwise specified)	C54.0–54.9, C55.9	8000–9049, 9056–9139, 9141–9589	Restrict to females
Ovary	C56.9	8000–9049, 9056–9139, 9141–9589	Restrict to females
Kidney	C64.9	8000–9049, 9056–9139, 9141–9589	
Meningioma	C70.0-70.1, 70.9	9530-9539	
Thyroid	C73.9	8000–9049, 9056–9139, 9141–9589	

Physical inactivity-associated cancers[5,12]

Cancer	ICD-O-3 site codes	ICD-O-3 histology codes	Additional restrictions
Colon	C18.0–18.9, C26.0	8000–9049, 9056–9139, 9141–9589	
Postmenopausal female breast	C50.0–50.9	8000–9049, 9056–9139, 9141–9589	Restrict to females and restrict to age ≥50 years
Corpus and uterus, NOS (not otherwise specified)	C54.0–54.9, C55.9	8000–9049, 9056–9139, 9141–9589	Restrict to females

Tobacco-associated cancers[6]

Cancer	ICD-O-3 site codes	ICD-O-3 histology codes	Additional restrictions
Lip, oral cavity, pharynx	C00.0–14.8	8000–9049, 9056–9139, 9141–9589	
Esophagus	C15.0–15.9	8000–9049, 9056–9139, 9141–9589	
Stomach	C16.0–16.9	8000–9049, 9056–9139, 9141–9589	
Colon and rectum	C18.0–20.9, C26.0	8000–9049, 9056–9139, 9141–9589	
Liver	C22.0	8000–9049, 9056–9139, 9141–9589	
Pancreas	C25.0–25.9	8000–9049, 9056–9139, 9141–9589	
Larynx	C32.0–32.9	8000–9049, 9056–9139, 9141–9589	
Trachea, lung, bronchus	C33.9–34.9	8000–9049, 9056–9139, 9141–9589	
Cervix uteri	C53.0–53.9	8000–9049, 9056–9139, 9141–9589	Restrict to females
Kidney and renal pelvis	C64.9–65.9	8000–9049, 9056–9139, 9141–9589	
Urinary bladder	C67.0–67.9	8000–9049, 9056–9139, 9141–9589	
Acute myeloid leukemia		9840; 9861; 9865–9869; 9871–9874; 9877–9879; 9910–9912; 9920	

References

1. Henley SJ, Singh SD, King J, Wilson RJ, O'Neil ME, Ryerson AB. Invasive cancer incidence and survival—United States, 2013. *MMWR Mortal Wkly Rep.* 2017;66(3):69–75.

2. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans: Volume 100E: Personal Habits and Indoor Combustions: Consumption of Alcoholic Beverages. Lyon, France: *International Agency for Research on Cancer* 2012;100E:373–499.

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

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

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

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

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

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

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

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

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

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

Sample Text for Manuscripts

Methods

Cancer incidence

U.S. Cancer Statistics data, which combine cancer registry data from the Centers for Disease Control and Prevention's (CDC's) National Program of Cancer Registries (NPCR) and the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) Program, were analyzed.[1] This dataset includes cancer incidence data from central cancer registries reported to NPCR in 46 states, the District of Columbia, **[IF APPLICABLE]**[2] and SEER in 4 states. Data about all new diagnoses of cancer from patient records at medical facilities such as hospitals, physicians' offices, therapeutic radiation facilities, freestanding surgical centers, and pathology laboratories are reported to central cancer registries, which collate these data and use state vital records to collect information about any cancer deaths that were not reported as cases. The central cancer registries use uniform data items and codes as documented by the North American Association of Central Cancer Registries. These data are submitted annually to CDC and NCI and combined into one dataset.[3] Cancer registries demonstrate that data are of high quality by meeting U.S. Cancer Statistics publication criteria;[1] during **[YEARX–YEARY]**, data from **[X]** cancer registries met these criteria, covering **[X%]** of the United States population. This report includes new cases of primary invasive **[CANCER TYPE]** cancer (International Classification of Diseases for Oncology, Third Edition code **[CXX.X–CXX.X]**)[4] diagnosed during **[YEARX–YEARY]**; **[IF APPLICABLE]** excluding histology codes 9050–9055, 9140, and 9590–9992 **[OR]** restricted to histology codes **[XXXX–XXXX]**.

[IF APPLICABLE] Race and ethnicity

Data were analyzed by five major racial/ethnic groups: White, Black, American Indian and Alaska Native (AI/AN), Asian/Pacific Islander (API), and Hispanic. Information about race and Hispanic ethnicity were collected separately. An algorithm was applied to Hispanic ethnicity data to reduce misclassification of Hispanic persons as being of unknown ethnicity.[5] To reduce misclassification of AI/AN race, some central cancer registries link case data with the Indian Health Service (IHS) patient registration database, which contains records of individuals who are members of federally recognized tribes; cases linked with the IHS database were coded as AI/AN.[6]

Because states can opt not to present state-specific counts and rates for **[AS APPLICABLE: API, Hispanic, and AI/AN populations]**, these data are not shown for the following states **[CHECK STATE LIST at www.cdc.gov/united-states-cancer-statistics/public-use/cautionary-notes.html; FOR EXAMPLE, Because states can opt not to present state-specific counts and rates for AI/AN populations, these data are not shown for Illinois, Kansas, New Jersey, and New York.]**

[IF APPLICABLE] Histology

Analyses by histology included only cases that were microscopically confirmed (**[X%]** of cases).

[IF APPLICABLE] Stage

Stage is classified using a merged variable that spans the time periods when three different staging schemes were used: SEER Summary Stage 2000, Derived Summary Stage, and Summary Stage 2018. The staging criteria characterize cancers as localized, regional, distant, or unknown stage. Localized cancer is confined to the primary site; regional cancer has spread directly beyond the primary site (regional extension) or to regional lymph nodes; and distant cancer has spread to other organs (distant extension) or remote lymph nodes.[7]

Population estimates

Population estimates for rate denominators were a modification of annual county population estimates by age, sex, bridged race, and ethnicity produced by the U.S. Census Bureau in collaboration with CDC and with support from NCI.[8] Modifications incorporated bridged, single-race estimates that were derived from multiple-race categories in the Census

and accounted for known issues in certain counties. The modified county-level population estimates, summed to the state and national levels, were used as denominators in rate calculations.[8]

Statistical analysis

Incidence and death rates

Average annual rates for **[YEARX–YEARY]** per 100,000 population were age-adjusted (using 19 age groups) by the direct method to the 2000 U.S. standard population.[9] Corresponding 95% confidence intervals (CIs) were calculated as modified gamma intervals.[10] Rates based on fewer than 16 cases tend to have poor reliability and were not presented. To determine differences between subgroups, rate ratios were calculated; rates were considered statistically different if the 95% CIs of the rate ratios excluded 1.[11] Rates were calculated using SEER*Stat software version **[X.X.X]**.[12]

[IF APPLICABLE] trends in rates

Annual percentage change (APC) was used to quantify the change in rates during **[YEARX–YEARY]** and was calculated using weighted least squares regression.[13] A two-sided t-test was used to test whether the APC was statistically different from zero (P <.05). Rates were considered to increase or decrease if P <.05; otherwise rates were considered stable. APCs were calculated using SEER*Stat software version **[X.X.X]**.[12]
[OR]
Change in rates during **[YEARX–YEARY]** was calculated using joinpoint regression, which involves fitting a series of joined straight lines on a logarithmic scale to the trends in the annual age-standardized rates;[14] up to **[X]** joinpoints (**[X]** line segments) were allowed. The trend of the line segment was used to quantify the annual percentage change (APC). A two-sided t-test was used to test whether the APC was statistically different from zero (P <.05). The average annual percentage change (AAPC) for **[YEARX–YEARY]** was calculated using a weighted average of the slope coefficients of the underlying joinpoint regression line with the weights equal to the length of each segment over the interval. To determine whether the AAPC was statistically different from zero (P <.05), a two-sided t-test was used for 0 joinpoints, and a two-sided z-test was used for 1 or more joinpoints. Rates were considered to increase or decrease if P <.05; otherwise rates were considered stable. Trends were calculated using Joinpoint regression program version **[X.X.X]**.[15]

Footnotes for tables

It is recommended that standard footnotes from U.S. Cancer Statistics or slight derivations be used for tables and figures.

For population coverage

Data are from population-based registries that participate in CDC's National Program of Cancer Registries and/or NCI's Surveillance, Epidemiology, and End Results Program and meet high-quality data criteria. These registries cover approximately **[XX]**% of the United States population.

For age-adjusted rates

Rates are per 100,000 persons and are age-adjusted to the 2000 U.S. standard population (19 age groups – Census P25–1130).

References

1. Centers for Disease Control and Prevention. U.S. Cancer Statistics. www.cdc.gov/united-states-cancer-statistics/ Accessed [ENTER DATE].
2. Singh SD, Henley SJ, Ryerson AB. Surveillance for cancer incidence and mortality—United States, 2012. *MMWR Morb Mortal Wkly Rep.* 2016;63(55):17–58.
3. National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: [ENTER DATABASE TITLE], United States Department of Health and Human Services, Centers for Disease Control

and Prevention. Released [DATE], based on the November [YEAR] submissions. Available at www.cdc.gov/united-states-cancer-statistics/public-use/.

4. Fritz A, Percy C, Jack A, Shanmugarathnam K, Sobin L, Parkin D, et al., editors. *International Classification of Diseases for Oncology. 3rd edition.* Geneva, Switzerland: World Health Organization; 2000.

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

6. Jim MA, Arias E, Seneca DS, et al. Racial misclassification of American Indians and Alaska Natives by Indian Health Service Contract Health Service Delivery Area. *Am J Public Health.* 2014;104:S295–S302.

7. Young JL Jr, Roffers SD, Ries LAG, Fritz AG, AA H, editors. SEER Summary Staging Manual – 2000: Codes and Coding Instructions. Bethesda, MD: National Cancer Institute; 2001.

8. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program. Modifications to Census Bureau's County Population Data.

9. Anderson R, Rosenberg H. Age standardization of death rates: implementation of the year 2000 standard. *Natl Vital Stat Rep.* 1998;47:1–16.

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

11. Fay MP. Approximate confidence intervals for rate ratios from directly standardized rates with sparse data. *Communications in Statistics: Theory and Methods* 2007;28(9):2141–2160.

12. National Cancer Institute. SEER*Stat software. Bethesda, MD: National Cancer Institute, Surveillance Research Program; [YEAR]. Available at https://seer.cancer.gov/seerstat/.

13. Kleinbaum DG, Kupper LL, Muller KE. *Applied Regression Analysis and Other Multivariable Methods.* 2nd ed. Boston, Mass: PWS-Kent; 1988.

14. Kim H-J, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat. Med.* 2000;19:335–351.

15. National Cancer Institute. Joinpoint Trend Analysis Software. Bethesda, MD: National Cancer Institute, Surveillance Research Program, Statistical Methodology and Applications Branch; [YEAR]. Available at https://surveillance.cancer.gov/joinpoint/.

Questions and Answers

I am having trouble using SEER*Stat. Where can I get help?

Step-by-step instructions for using SEER*Stat are available at https://seer.cancer.gov/seerstat/tutorials. If you do not find answers there, please contact SEER*Stat Technical Support at seerstat@imsweb.com.

Are county-level data available?

Due to data-sharing agreements with some of the states providing data for this database, county-level data are not available in the public use database. County-level data are available to researchers in the U.S. Cancer Statistics Restricted Access Database, which is hosted at CDC's National Center for Health Statistics Research Data Center.

Are single years of age available?

Single-year age data are not available in the public use database. Single-year age data are available to researchers in the U.S. Cancer Statistics Restricted Access Database.

How do I obtain mortality data?

When you obtain access to SEER Research Plus, CDC's National Health Statistics' U.S. Mortality data files are included in your SEER*Stat profile. Please refer to SEER's U.S. Mortality Data website at https://seer.cancer.gov/mortality/ for information on how to use these data. Mortality data are also available in the U.S. Cancer Statistics Data Visualizations tool at https://www.cdc.gov/cancer/dataviz and CDC WONDER at https://wonder.cdc.gov/cancer.html.

Can I link this database to other data sources?

The U.S. Cancer Statistics Database Request Form prohibits analysts from linking this database with any other database without CDC's written approval. Please contact uscsdata@cdc.gov if you wish to link the data.

Is a SAS or ASCII dataset available?

No. The data can be analyzed only within the SEER*Stat software.

Is case listing available? Can I download the data?

Due to data-sharing agreements with some of the states providing data for this database, case listing is not available in the SEER*Stat software, and case-level data cannot be accessed or downloaded.