U.S. Cancer Statistics Public Use Database Technical Documentation

U.S. Data
November 2022 Submission
Diagnosis Years 2001–2021

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, 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 37 million cancer cases. 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, providing information on more than 37 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 2023 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, 2021. 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: 37,277,847 (includes benign and borderline brain and other nervous system tumors from 2004 onward)
- Malignant cases†: 33,622,934
- Malignant and *in situ* cases†: 36,234,867

*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. These registries cover approximately **[XX]**% 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 2001–2021 Public Use Research Database, 2023 submission (2001–2021), U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June 2024. 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 22 U.S. geographic areas.

Central cancer registries submitting data to CDC and NCI during 2023 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	

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 with sex other than male or female 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.

The 2023 data submission includes new cancer cases diagnosed in 2020 and 2021, the first and second years of the COVID-19 pandemic. The COVID-19 pandemic disrupted health services, leading to delays and reductions in cancer screening and diagnosis, which may have contributed to lower incidence for most cancer sites in 2020. The number of new cases diagnosed in 2021 are still a little lower for some cancer types but have returned to pre-pandemic counts for other cancer types.[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.[2] JoinPoint software allows researchers to exclude incidence data for 2020, 2021, or both years from trend analyses. Exclude 2020 data for incidence trend analyses, but 2021 data can be included in incidence trend analyses.

Suppression rules

Suppressing fewer than 16 cases

The suppression rule[3] [4] 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 counts, rates, and trends that are unstable because of the small number of cases, these statistics are not shown in tables and figures if the counts are fewer than 16 for the time period. A count of fewer than about 16 in a numerator results in a standard error of the rate that is about 25% or more as large as the rate itself. Equivalently, a count of fewer than about 16 results in the width of the rate's 95% confidence interval being at least as large as the rate itself. These relationships were derived under the assumption of a Poisson process and with the standard population age distribution 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 subtracting to find suppressed counts. Use this practice when any suppression occurs in the data presentation. In addition, when information from other cells, tables, or figures can be used to determine a suppressed cell, suppress counts and rates for at least one other cell. 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.

Benign 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.[5 6 7 8 9]

Consider reviewing the variable *Site recode ICD-O-3/WHO 2008* before using the directly coded primary site.

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.[10] 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; 2023.

2. Surveillance, Epidemiology, and End Results Program. Impact of COVID on the April 2024 SEER Data Release. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2024.

3. Federal Committee on Statistical Methodology. *Report on Statistical Disclosure Limitations Methodology (Statistical Working Paper 22).* Washington, DC: Office of Management and Budget; 2005.

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

5. Fritz A, Percy C, Jack A, et al., editors. *International Classification of Diseases for Oncology, Third Edition.* Geneva: World Health Organization; 2000.

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

7. Ruhl J, Adamo M, Dickie L., Negoita, S. (September 2020). Hematopoietic and Lymphoid Neoplasm Coding Manual. National Cancer Institute, Bethesda, MD, 2020.

8. Surveillance, Epidemiology, and End Results Program. 2024 Solid Tumor Rules. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2023.

9. Surveillance, Epidemiology, and End Results Program. Hematopoietic and Lymphoid Neoplasm Database. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2016.

10. 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 2001–2021:

- 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 2022 postcensal estimates (for July 1, 2020–2022) 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 2017, 100% of the population is covered for all 50 states and the District of Columbia.
- In 2020 and 2021, cases that were diagnosed in Indiana are not available. The U.S. population covered for 2020 and 2021 is 98.0%.
- The U.S. population coverage for 2001 through 2021 is 99.6%.

U.S. and Puerto Rico public use database

A database including U.S. and Puerto Rico data is no longer released publicly. 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*.

Change History

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 *USCS0121* for diagnosis years 2001–2021).
- The most recent 10 years of data (*USCS1221* for diagnosis years 2012–2021).
- The most recent 5 years of data (*USCS1721* for diagnosis years 2017–2021).

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 *USCS0121, USCS1221,* or *USCS1721* 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 *USCS1721* is used, then *Year of Diagnosis* should also be restricted to diagnosis years 2017–2021 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 2001–2021 Public Use Research Database?

Variable Definitions

The following variables are available in the U.S. Cancer Statistics Public Use Database, U.S. data (2001–2021). They are listed by SEER*Stat category.

Age at Diagnosis

- Age recode with <1 year olds

Race, sex, year of diagnosis, and registry

- Sex
- Year of Diagnosis
- Addr at Dx -state
- USCS Standard
- Race recode (W, B, AIAN, API)
- Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)
- Program
- Region
- USCS0121
- USCS1221
- USCS1721
- Origin recode NHIA (Hispanic, non-Hisp)

Site and morphology

- Primary site -labeled
- Histologic type ICD-O-3 *(International Classification of Diseases for Oncology, Third Edition)*
- Behavior code ICD-O-3
- Grade
- Grade clinical
- Grade pathological
- Diagnostic conformation
- ICD-O-3 histology/behavior, labeled
- Laterality
- Site recode ICD-O-3/WHO 2008
- Schema ID
- ICCC recode 3rd edition ICD-O-3/IARC 2017
- ICCC recode extended 3rd edition ICD-O-3/IARC 2017
- AYA site recode 2020 revision
- Lymphoid neoplasm recode 2021 revision

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

- Merged summary stage

Therapy

- Rx summary – surgery primary site

Extent of disease

- CS site-specific factor 1
- Merged estrogen receptor
- Merged progesterone receptor
- Merged HER2 summary

Multiple primary fields

- Sequence number – central

Dates

- Year of birth
- Month of diagnosis

User-specified

- Rural-urban continuum 2013, grouped

Merged system-supplied

- Alcohol-related cancers
- Human papillomavirus (HPV) -related cancers
- Obesity-related cancers
- Physical inactivity-related cancers
- Tobacco-related cancers
- State race ethnicity suppress