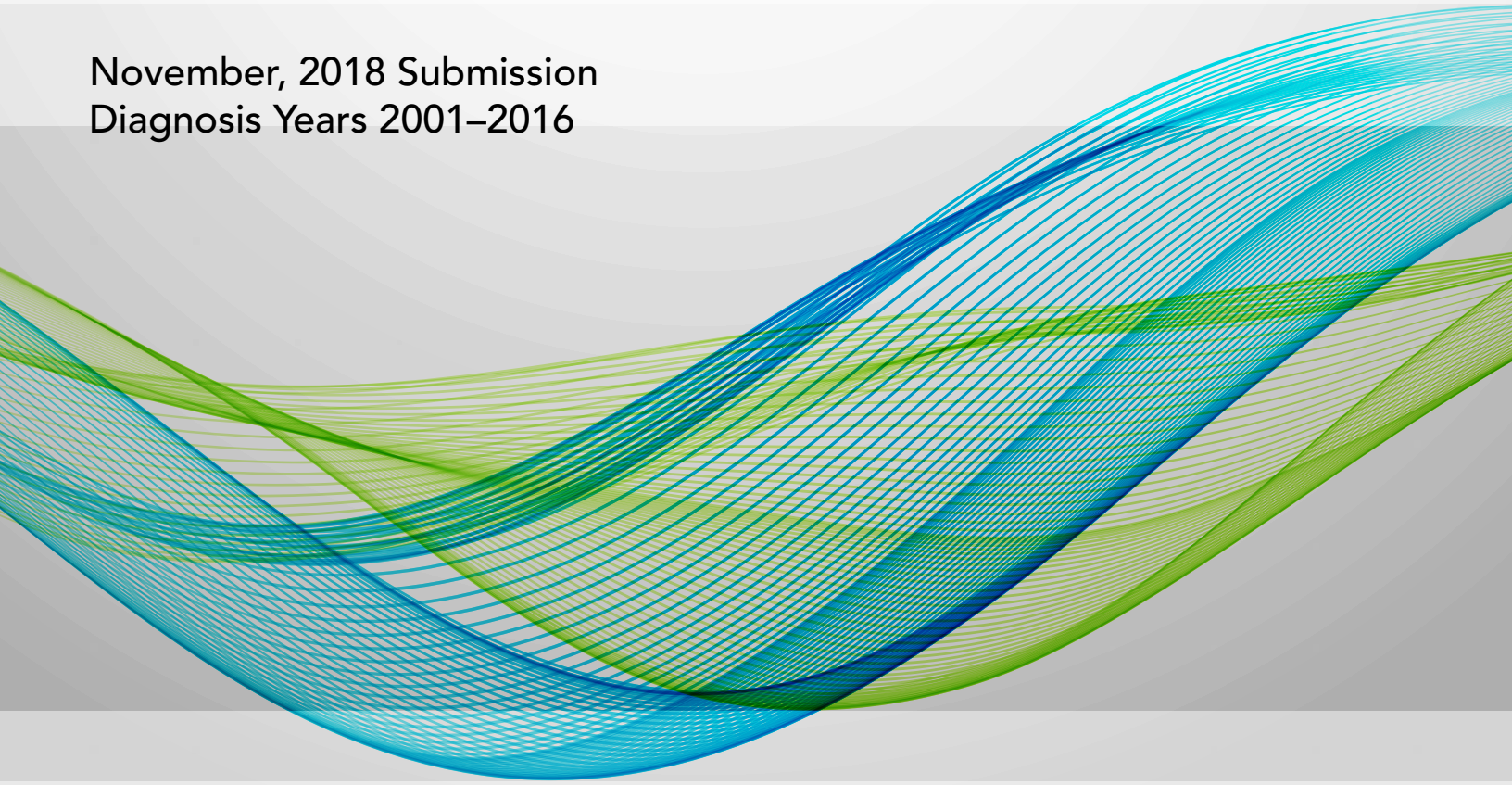
National Program of Cancer Registries (NPCR) and Surveillance, Epidemiology & End Results (SEER)

NPCR and SEER Incidence—U.S. Cancer Statistics 2001–2016 Public Use Database Data Standards and Data Dictionary

November, 2018 Submission
Diagnosis Years 2001–2016



U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

NIH NATIONAL CANCER INSTITUTE

CS304502-A

Table of Contents

Message to Data Users

June 6, 2019

We are happy to share the 2019-release of the U.S. Cancer Statistics public use dataset from CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. This database provides population-based cancer statistics on the *entire* United States population.

The NPCR and SEER Program are comprehensive cancer surveillance systems that work collaboratively with partners to collect, compile, and disseminate information on more than 1.8 million cancer cases annually. U.S. Cancer Statistics data products, like this public use database, are made possible by the dedicated efforts of reporting facilities, cancer registrars, central cancer registries, and CDC NPCR and NCI SEER staff and contractors. I thank everyone for their contributions in collecting these important and high quality data.

Cancer registry data provide a foundation of cancer surveillance activities that are used to measure progress and target cancer prevention and control activities. We encourage researchers to use our data to inform scientific inquiries, programs, and policies. Through use of this U.S. Cancer Statistics public use data source, researchers can positively impact comprehensive cancer prevention and control as well as the care and quality of lives for those diagnosed with cancer.

Sincerely,

Vicki Benard, PhD
Branch Chief, Cancer Surveillance Branch
Division of Cancer Prevention and Control
National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention

Overview of CDC's National Program of Cancer Registries and NCI's Surveillance, Epidemiology, and End Results Program

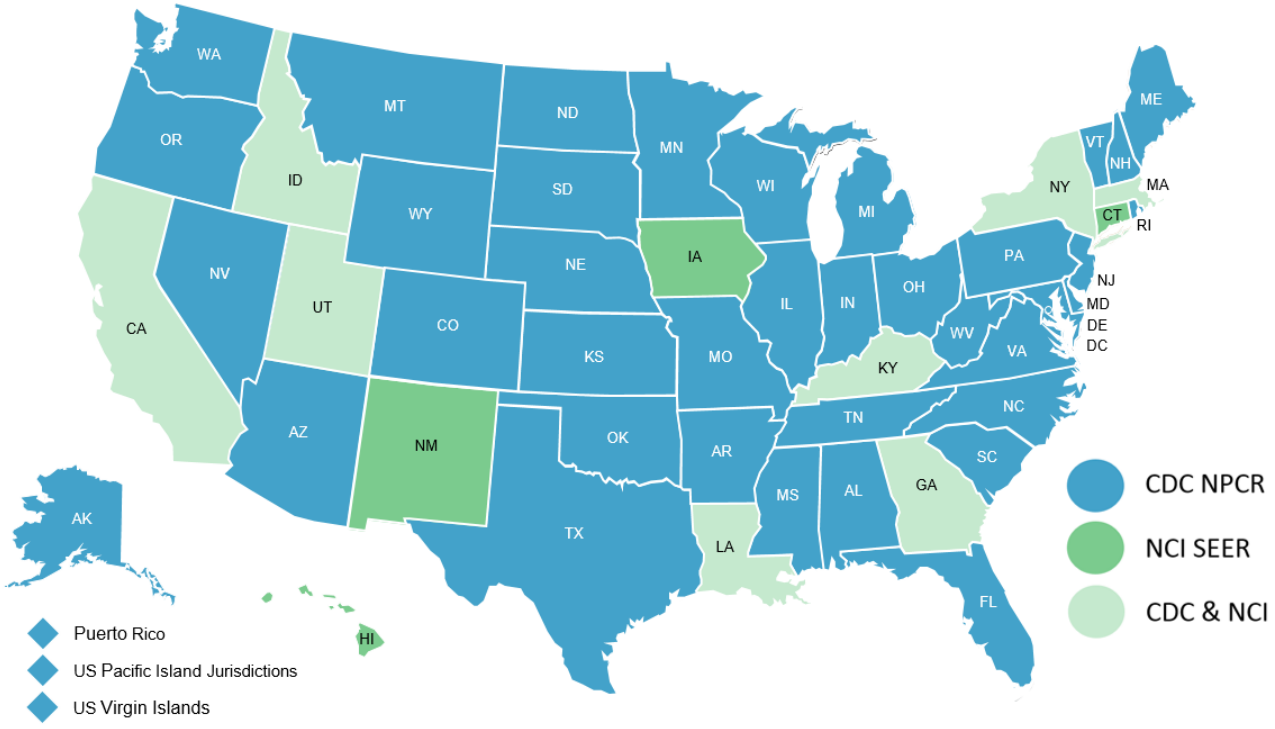
The National Program of Cancer Registries (NPCR), administered by the Centers for Disease Control and Prevention (CDC), was established by Congress in 1992. Through cooperative agreements, NPCR supports central cancer registries in 46 states, the District of Columbia, Puerto Rico, U.S. Pacific Island Jurisdictions, and the U.S. Virgin Islands (see Figure 1).

The Surveillance, Epidemiology, and End Results (SEER) Program, administered by the National Cancer Institute (NCI), has been funded since 1973 as a result of the National Cancer Act of 1971. SEER collects reportable cancer cases from 19 U.S. geographic areas, including 12 states (see Figure 1 below). Together, CDC's NPCR and NCI's SEER Program cover the entire United States population. These combined data are the official source of federal statistics on cancer incidence and are referred to as the U.S Cancer Statistics.

The cancer registries funded by CDC and NCI routinely collect data on patient demographics, primary tumor site, tumor morphology, stage at diagnosis, first course of treatment, and outcomes. Medical facilities such as hospitals, doctor's offices, pathology laboratories, and other treatment centers send demographic and clinical information related to people diagnosed with cancer to a central cancer registry, where the information is consolidated and goes through rigorous checks for quality and completeness. All hospitals are required by state law to report cancer cases to the central cancer registry in their respective states. On an annual basis, the central cancer registries submit demographic and clinical information about each cancer case to CDC and/or NCI.

This national coverage of cancer data from CDC's NPCR and NCI's SEER Program enables researchers, clinicians, policy makers, public health professionals, and members of the public to monitor the burden of cancer, evaluate the success of programs, and identify additional needs for cancer prevention and control efforts at national, state, and local levels.

Figure 1. Central cancer registry programs submitting data to NPCR and SEER in 2018



Data Collection, Submission, Quality, and Coverage

The most current data come from the November 2018 NPCR and SEER submissions, which include cancer cases diagnosed from January 1, 2001 through December 31, 2016. Each year, NPCR- and SEER-funded central cancer registries submit data on cancer diagnosed during the most recent year to the respective program. In addition, data from previous years are updated with information from the newly submitted records to ensure case completeness and high quality. NPCR and SEER allow an interval of 23 months after the close of the diagnosis year for submission (for the 2016 data, NPCR required submission by November 30, 2018 and SEER required submission by November 1, 2018).

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 on all NPCR and SEER data. For a registry's data to be included in the U.S. Cancer Statistics public research data file, they must have met the following quality and completeness criteria for publication[1]—

- 5% or fewer cases are ascertained solely on the basis of a death certificate.
- 3% or fewer cases are missing information on sex.
- 3% or fewer cases are missing information on age.
- 5% or fewer cases are missing information on race.
- 97% or more of the registry's records passed a set of single-field and interfield computerized edits.

NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 Public Use Database

Two NPCR and SEER Incidence – U.S. Cancer Statistics public use databases are available for researchers: the 2001–2016 database and the 2005–2016 database. **This data standards document is specific to the 2001–2016 database**.

The 2001–2016 database includes race and ethnicity variables, while the 2005–2016 database does not. The 2005–2016 database includes Puerto Rico data, while the 2001–2016 database does not.

- The 2001–2016 database's population denominators are race-specific, ethnicity-specific, and sex-specific county population estimates from the U.S. Census (July 1, 2010–2017 bridged–race vintage 2017 population estimates), modified by SEER and aggregated to the state and national levels.

- The 2005–2016 database's population denominators are sex-specific, are from the 2010 U.S. Census, and are not available by race or ethnicity.

Table 1 lists the population coverage by diagnosis year for the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 public research data. In the 2019 release of the public use database there is 100% population coverage for all 50 states and the District of Columbia for cases diagnosed from 2003 through 2016. Mississippi's cases diagnosed in 2001 and 2002 are not available and the U.S. population coverage those 2 years is 99%.

[1] Additional information is available at https://www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm

Table 1. U.S. population coverage [a], NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 Public Use Research Database.

Diagnosis year(s)	Percentage of U.S. population covered in database
2001 [b]	99%
2002 [b]	99%
2003	100%
2004	100%
2005	100%
2006	100%
2007	100%
2008	100%
2009	100%
2010	100%
2011	100%
2012	100%
2013	100%
2014	100%
2015	100%
2016	100%
2001–2016	99%
2007–2016 [c]	100%
2012–2016 [d]	100%

[a] The NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 Public Use Research Database includes data submitted by all 50 states and District of Columbia. Puerto Rico, U.S. Pacific Island Jurisdiction, and U.S. Virgin Island data are not included in the database.

[b] Mississippi's cases diagnosed in 2001 and 2002 are not available in the 2001-2016 public use database.

[c] The most recently submitted 10 years of data.

[d] The most recently submitted 5 years of data.

Variable List

Table 3 shows all of the variables available in the 2001-2016 public use research database.

Table 3. Variables in the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 Public Use Research Database

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
Age at Diagnosis	Age recode with <1 year olds	
Race, Sex, Year Dx, Registry, County	Sex	
	Year of diagnosis	
	Addr at DX – state	
	USCS standard	
	Race recode for USCS	
	Program	
	Region	
	USCS0116	
	USCS0716	
	USCS1216	
	Origin recode NHIA (Hispanic, Non-Hisp)	
Site and Morphology	Primary site – labeled	
	Histologic type ICD-O-3	
	Grade	
	Diagnostic confirmation	
	ICD-O-3 hist/behavior, labeled	
	Site recode ICD-O-3/WHO 2008	
	ICCC site recode ICD-O-3/WHO 2008	
	ICCC site rec extended ICD-O-3/WHO 2008	
	AYA site recode/WHO 2008	
	Lymphoma subtype recode/WHO 2008	
	Behavior recode for analysis derived/WHO2008	
Stage – LRD [Summary and Historic]	Merged summary stage 2000	
Therapy	RX summ – surg prim site	Female breast only and diagnosis years ≥2003
Extent of Disease – CS	CS site-specific factor 1	Restricted to 2 groups: - Female breast and diagnosis years ≥2004 - Brain and diagnosis years ≥2011
	CS site-specific factor 2	Female breast only and diagnosis years ≥2004
	CS site-specific factor 15	Female breast only and diagnosis years ≥2010
	Laterality	
Multiple Primary Fields	Sequence number – central	
Race and Age (case data only)	NHIA derived Hisp origin	
Dates	Year of birth	

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
	Month of diagnosis	
Other	Type of reporting source	
Merged System-Supplied	Alcohol-related cancers	
	HPV-related cancers	
	Obesity-related cancers	
	Physical inactivity-related cancers	
	Tobacco-related cancers	
	State race eth suppress	

Abbreviations used in the variable names –

Addr	Address
AYA	Adolescent and young adult
CS	Collaborative stage
Dx	Diagnosis
Hisp	Hispanic
ICCC	International Classification of Childhood Cancer
ICD-O-3	*International Classification of Diseases for Oncology*, Third Edition
LRD	Local, regional, distant
NHIA	NAACCR Hispanic identification algorithm
USCS	U.S. Cancer Statistics
WHO	World Health Organization

Data Citations

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

- **For population coverage[2]:** 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 2001–2016 database:** National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Database, Nov 2018 submission (2001-2016), United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June 2019, based on November 2018 submissions. Available at www.cdc.gov/cancer/public-use.

[2] See Table 1 for percentage population coverage applicable to years being analyzed.

Cautionary Notes

Before using this database, analysts should read and understand the following section. If you have questions regarding these notes, please contact CDC at uscsdata@cdc.gov.

Case Inclusions and Exclusions

NPCR- and SEER-supported cancer registries report all incident cases coded as *in situ* (non malignant) and invasive (malignant; primary site only) 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.
- Non-malignant (including borderline and benign) central nervous system tumors are reported.
- *In situ* cancers of the urinary bladder are re-coded as invasive behavior because the information needed to distinguish 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 public use database –

- Cancer cases that were identified only through death certificate or autopsy reports have been excluded.
- Cases with an unknown age or with sex other than male or female have been excluded from the database. The frequency counts presented in this document 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.

  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.

Suppression Rules[2-3]

Complementary Cell Suppression

When analyzing data at the state or regional levels, counts for national and regional data must be suppressed if a single state in a region or division is suppressed. This practice is referred to as *complementary cell suppression* and is necessary to prevent users from subtracting to find suppressed counts. Rates, confidence intervals (CIs), and populations can be shown at the national and regional levels. This suppression should occur when a single or multiple years of data are being presented.

Suppressing fewer than 16 cases

The suppression rule is fewer than 16 cases for the time period based on rate stability. This suppression rule is enforced automatically in this database.

When the numbers of cases used to compute the incidence rates are 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 95% confidence interval around the rate 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.

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 eth 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 state-level race or ethnicity data presentation restrictions:

- Data for American Indians and Alaska Natives cannot be displayed for Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York.
- Data for Asians and Pacific Islanders cannot be displayed for Delaware, Kansas and Kentucky.
- Hispanic ethnicity data cannot be displayed for Delaware, Kentucky, and Massachusetts.
- Race and ethnicity combinations—e.g., white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.

For more information, please refer to the *Race Recode for USCS*, *Origin recode NHIA (Hispanic, Non-Hisp)*, and *NHIA derived Hisp origin* variable descriptions in this document.

Case Level Data

As a further mechanism to protect data confidentiality and due to data sharing agreements with some of the states providing data for this database, 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) and those with a behavior code of 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.

Primary Site Variables[4]

Beginning in diagnosis year 2010, some of the lymphoma and leukemia ICD-O-3 codes were updated based on changes from the World Health Organization. The appropriate site recode variables to include these updates are *Site recode ICD-O-3/WHO 2008* for all ages and *International Classification of Childhood Cancer (ICCC) site recode ICD-O-3/WHO 2008* and *ICCC site rec extended ICD-O-3/WHO 2008* for the childhood cancer recodes.

Consider reviewing the variable *Site recode ICD-O-3/WHO 2008* before using the directly coded primary site. For more information on the SEER primary site recodes, see http://seer.cancer.gov/siterecode.

Histologic Type ICD-O-3[5-8]

Beginning with 2010 diagnoses, this item also includes histology codes as specified in the 2008 World Health Organization (WHO) hematopoietic/lymphoid publication, which are listed on pages 3–5 of the NAACCR *2010 Implementation Guidelines and Recommendations*, available at www.facs.org/~/media/files/quality%20programs/cancer/coc/2010implementationguidelines.ashx.

Stage

A merged variable, *Merged Summary Stage 2000*, has been created to span three time periods when two different staging schemes were used. Stage at diagnosis is recorded using *SEER Summary Stage 2000* for diagnosis years 2001–2003, *Derived SEER Summary Stage 2000* for diagnosis years 2004–2015, and *SEER Summary Stage 2000* for diagnosis year 2016.

The coding logic for this merged variable is:

- If a case was diagnosed between 2001 and 2003, stage at diagnosis is recorded using the *SEER Summary Stage 2000* variable value.
- If a case was diagnosed between 2004 and 2015, then the stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value.
- If a case was diagnosed in 2016, then the stage at diagnosis is recorded using the *SEER Summary Stage 2000*.
- If the *Derived SEER Summary Stage 2000* variable is blank and a valid value is available for the *SEER Summary Stage 2000* variable, that value is used to populate the merged variable. For example, if a case was diagnosed in 2015 and *Derived SEER Summary Stage* was blank, but *SEER Summary Stage* had a value of *local*, then the merged variable was coded as local stage. Otherwise, the merged variable left blank.

Reporting Delay[9]

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.* 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.

References

1. Young JL Jr, Roffers SD, Ries LAG, Fritz AG, Hurlbut AA (eds). *SEER Summary Staging Manual - 2000: Codes and Coding Instructions*, National Cancer Institute, NIH Pub. No. 01-4969, Bethesda, MD, 2001.
2. Federal Committee on Statistical Methodology. *Report on Statistical Disclosure Limitations Methodology (Statistical Working Paper 22).* Washington, DC: Office of Management and Budget; 2005. Available at https://fcsm.sites.usa.gov/files/2014/04/spwp22.pdf.
3. Doyle P, Lane JI, Theeuwes JM, Zayatz LM. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies.* Amsterdam: Elsevier Science; 2001.
4. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin D, et al., editors. *International Classification of Diseases for Oncology.* Third Edition. Geneva: World Health Organization; 2000.
5. *International Classification of Diseases for Oncology.* Third Edition, First Revision. Geneva: World Health Organization, 2013.
6. Ruhl J, Adamo M, Dickie L. (January 2015). *Hematopoietic and Lymphoid Neoplasm Coding Manual*. National Cancer Institute, Bethesda, MD 20850-9765.
7. Surveillance, Epidemiology, and End Results Program. *2007 Multiple Primary and Histology Coding Rules*. Bethesda, MD: US Department of Health and Human Services, National Cancer Institute; Revised August 24, 2012; Accessed January 25, 2017. https://seer.cancer.gov/tools/mphrules/.

8. Surveillance, Epidemiology, and End Results Program. *Hematopoietic and Lymphoid Neoplasm Database*. Bethesda, MD: US Department of Health and Human Services, National Cancer Institute; 2016. https://seer.cancer.gov/seertools/hemelymph.

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

Checklist for a NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 Public Use Data Analysis

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 *USCS0116* for diagnosis years 2001–2016), the most recent 10 years of data (*USCS0716* for diagnosis years 2007–2016) and the most recent 5 years of data (*USCS1216* for diagnosis years 2012–2016).

If you are conducting a multi-year analysis and want to restrict it to the states that met reporting standards during each of the years, did you use variable *USCS0116*, *USCS0716*, or *USCS1216* and also use the *Year of Diagnosis* variable to restrict to the corresponding year range on the SEER*Stat Selection tab?

- This is important for trend analyses, as the same states need to be included for each year being analyzed for comparable comparisons.
- The *Year of Diagnosis* variable is used in combination with the predefined USCS variable to exclude the non-relevant years. For example, if *USCS1216* is used, then *Year of Diagnosis* should also be restricted to diagnosis years 2012–2016 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.[1]

Single year analyses

☐ If you are analyzing just 1 year of data, did you use the variable *USCS Standard* and restricted the analysis to the specific *Year of Diagnosis* in the SEER*Stat Selection tab? [2]

Common selection and reporting considerations

☐ 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 eth suppress* variable; race-only or ethnicity–only suppressions be done manually in the SEER*Stat Selection tab.[3]

☐ 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)?[4]

☐ 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?[5]

☐ 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?[6]

☐ When reporting **rates**, have you included the label “per 100,000 persons,” “per 100,000 women,” or “per 100,000 men”?

☐ Have you included **citations** for the:

- Percentage of United States population coverage provided by the database?
- NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 Public Use Research Database?[7]

---

[1] See *USCS0116*, *USCS0716*, and *USCS1216* variable descriptions.

[2] See *USCS Standard* variable description.

[3] See *state race eth suppress*, *Race Recode*, *Origin recode NHIA*, and *NHIA derived Hisp origin* variable descriptions.

[4] See Cautionary Notes section entitled *Primary Site Variables*.

[5] See *Diagnostic Confirmation* variable descriptions.

[6] See *Sex* variable description.

[7] See *Data Citation* section.

NPCR and SEER Incidence – U.S. Cancer Statistics 2001– 2016 Public Use Database: Variable Definition and Frequency

Please note that the frequencies presented in this document were created with *Malignant Behavior* unselected on the SEER*Stat Selection tab.

SEER*Stat Item Name: **Age recode with <1 year olds**
Source of Standard: NAACCR
Source Item Name: Derived from *Age at diagnosis*
Source Item Number: 230

Description

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	Frequency	Percentage
00 years	18,273	0.1%
01–04 years	59,820	0.2%
05–09 years	44,707	0.2%
10–14 years	52,793	0.2%
15–19 years	89,076	0.3%
20–24 years	142,376	0.5%
25–29 years	225,160	0.8%
30–34 years	347,724	1.3%
35–39 years	531,996	2.0%
40–44 years	919,632	3.4%
45–49 years	1,476,443	5.5%
50–54 years	2,220,186	8.3%
55–59 years	2,883,111	10.8%
60–64 years	3,360,959	12.6%
65–69 years	3,683,013	13.8%
70–74 years	3,412,871	12.8%
75–79 years	3,018,065	11.3%
80–84 years	2,281,385	8.6%
≥85 years	1,890,480	7.1%

SEER*Stat Item Name: **Sex**

Source of Standard: NAACCR
Source Item Name: *Sex*
Source Item Number: 220

Description

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 and the lack of an associated population file, cases for sex other than male or female are excluded from this database.

Values	Frequency	Percentage
Male	13,214,617	49.6%
Female	13,443,453	50.4%

SEER*Stat Name: **Year of diagnosis**
Source of Standard: NAACCR
Source Item Name: Derived from *Date of initial diagnosis (CoC)*
Source Item Number: 390

Description

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.
- For more information, please see
  - NAACCR data dictionary https://www.naaccr.org/data-standards-data-dictionary
  - FORDS www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals
  - SEER coding manuals http://seer.cancer.gov/tools/codingmanuals/historical.html

Values	Frequency	Percentage
2001	1,441,526	5.4%
2002	1,461,749	5.5%
2003	1,474,784	5.5%
2004	1,537,628	5.8%
2005	1,572,851	5.9%
2006	1,613,076	6.1%
2007	1,666,822	6.3%
2008	1,691,353	6.3%
2009	1,718,114	6.4%
2010	1,709,157	6.4%
2011	1,754,818	6.6%
2012	1,750,091	6.6%
2013	1,783,334	6.7%
2014	1,807,920	6.8%
2015	1,845,808	6.9%
2016	1,829,039	6.9%

SEER*Stat Item Name: **Addr at DX – State**
Source of Standard: NAACCR
Source Item Name: *State at diagnosis (CoC)*
Source Item Number: 80

Description

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.
- For more information, please see
  - NAACCR data dictionary https://www.naaccr.org/data-standards-data-dictionary
  - FORDS variable "state at diagnosis" at www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals

Values	Frequency	Percentage
Alaska	45,289	0.2%
Alabama	417,733	1.6%
Arkansas	255,390	1.0%
Arizona	472,218	1.8%
California	2,756,513	10.3%
Colorado	363,090	1.4%
Connecticut	363,281	1.4%
District of Columbia	49,803	0.2%
Delaware	91,449	0.3%
Florida	1,965,511	7.4%
Georgia	747,205	2.8%
Hawaii	115,015	0.4%
Idaho	122,590	0.5%
Illinois	1,122,484	4.2%
Indiana	559,280	2.1%
Iowa	295,725	1.1%
Kansas	247,855	0.9%
Kentucky	430,062	1.6%
Louisiana	398,588	1.5%
Massachusetts	629,852	2.4%
Maryland	491,211	1.8%
Maine	145,518	0.5%
Michigan	933,827	3.5%
Minnesota	461,304	1.7%
Missouri	529,607	2.0%
Mississippi	223,975	0.8%
Montana	94,701	0.4%
North Carolina	830,121	3.1%
North Dakota	60,238	0.2%
Nebraska	158,501	0.6%

Values	Frequency	Percentage
New Hampshire	134,102	0.5%
New Jersey	870,218	3.3%
New Mexico	148,450	0.6%
Nevada	194,060	0.7%
New York	1,875,129	7.0%
Ohio	1,050,785	3.9%
Oklahoma	315,311	1.2%
Oregon	343,113	1.3%
Pennsylvania	1,332,614	5.0%
Rhode Island	106,985	0.4%
South Carolina	414,815	1.6%
South Dakota	72,982	0.3%
Tennessee	542,760	2.0%
Texas	1,679,866	6.3%
Utah	163,250	0.6%
Virginia	627,796	2.4%
Vermont	63,162	0.2%
Washington	592,003	2.2%
Wisconsin	515,943	1.9%
West Virginia	193,011	0.7%
Wyoming	43,779	0.2%

SEER*Stat Item Name: **USCS standard**

Source of Standard: NPCR

Source Item Name: Derived from NAACCR's *State at diagnosis (CoC)* and U.S. Cancer Statistics publication standard

Source Item Number: Derived from NAACCR's 80

Description

This variable indicates the central cancer registries with cancer incidence data that are of high quality and meet the U.S. Cancer Statistics standard for a single year of analysis at the national level for all cancer sites combined.

Considerations for use

- This variable allows the selection of only those central cancer registries whose data meet the U.S. Cancer Statistics standard for an individual diagnosis year. The year of diagnosis should also be specified in the SEER*Stat Selection tab.
- If you are conducting a multiyear analysis and want to restrict the analysis to the states that met publication criteria for each of those years (for example, a trend analysis), we recommend using the predefined variables, *USCS1216* (includes diagnosis years 2012–2016), *USCS0716* (includes diagnosis years 2007–2016), or *USCS0116* (includes diagnosis years 2001–2016).
- 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.
- The U.S. Cancer Statistics publication standard is available at www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

Number of central cancer registries[a]	Year of diagnosis	Frequency	Percentage
50	2001	1,441,526	5.4%
50	2002	1,461,749	5.5%
51	2003	1,474,784	5.5%
51	2004	1,537,628	5.8%
51	2005	1,572,851	5.9%
51	2006	1,613,076	6.1%
51	2007	1,666,822	6.3%
51	2008	1,691,353	6.3%
51	2009	1,718,114	6.4%
51	2010	1,709,157	6.4%
51	2011	1,754,818	6.6%
51	2012	1,750,091	6.6%
51	2013	1,783,334	6.7%
51	2014	1,807,920	6.8%
51	2015	1,845,808	6.9%
51	2016	1,829,039	6.9%

[a] Refer to Table 1 for the central cancer registries available by diagnosis year.

SEER*Stat Item Name: **Race Recode for USCS**

Source of Standard: NAACCR
Source Item Name: Derived from *Race 1*, *Race 2*, and *race- NAPIIA (derived API)*
Source Item Number: 160 (*Race 1*), 161 (*Race 2*), and 192 (*race- NAPIIA (derived API)*)

Description

This variable indicates the derived code for the patient's race. Race is coded separately from Hispanic ethnicity. This variable is created using NAACCR variables *Race1*, *Race2*, the Indian Health Service (IHS) Link variable, and *race-NAPIIA (derived API)*. Race recode starts as *Race1*. If *Race1* is white and *Race 2* is a specified non-white race, then the value from *Race2* is used. After this check, if Race is still white and there is a positive IHS link, then *Race/Ethnicity* is set to American Indian/Alaskan Native (AI/AN).

Considerations for use

- This variable is available only in the 2001–2016 public use database; it is not available in the 2005–2016 database.
- 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 American Indians and Alaska Natives cannot be displayed for Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York.
  - Data for Asians and Pacific Islanders cannot be displayed for Delaware, Kansas and Kentucky.
  - Data for race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
- 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.
- IHS provides medical services to American Indians and Alaska Natives (AI/ANs) who are members of federally recognized tribes, estimated to be about 65% of the AI/AN population. To improve identification of AI/ANs, 31 NPCR registries with Purchase/Referred Care Service Delivery Area (PRCSDA) counties in their state annually link with the IHS patient registration database to identify AI/ANs who had not been coded as AI/AN (shown in Appendix A). All NPCR registries link every five years; linkages were performed by all NPCR states most recently in 2016. In 2016, SEER registries linked cancer cases diagnosed from 1994–2014.
  - 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 restricting the NPCR states included in the analysis to NPCR registries that conduct annual IHS linkages. See Appendix A for the list of these states.
- In all separate records of tumors for the same patient, the patient should have the same race code.
- The *Race Recode for USCS* 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 https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity. Population data are not available for the "other race" and "unknown race" categories.
- For more information, please see
  - NAACCR data dictionary https://www.naaccr.org/data-standards-data-dictionary
  - FORDS at www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals
  - SEER coding manuals http://seer.cancer.gov/tools/codingmanuals/historical.html

Values		
White	22,679,353	85.1%
Black	2,745,850	10.3%
American Indian/Alaska Native	134,284	0.5%
Asian or Pacific Islander	730,342	2.7%
Other unspecified (1991+)	88,736	0.3%
Unknown	279,505	1.0%

SEER*Stat Item Name: **Program**

Source of Standard: NPCR
Source Item Name: Not applicable
Source Item Number: Not applicable

Description

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 NPCR and submitted any 2001–2016 diagnosis years data (i.e., 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) are categorized as "NPCR" states.
- "SEER" refers to central cancer registries receiving funding only from SEER during the 2001–2016 diagnosis years (i.e., Connecticut, Hawaii, Iowa, and New Mexico).

Values	Frequency	Percentage
NPCR	25,735,599	96.5%
SEER	922,471	3.5%

SEER*Stat Item Name: **Region**

Source of Standard: NPCR
Source Item Name: Derived from NAACCR's *State at diagnosis (CoC)* and US Census Region
Source Item Number: Derived from NAACCR's 80

Description

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 state registries that met U.S. Cancer Statistics publication criteria are included in the database. It is recommended 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 using the specified: www.cdc.gov/cancer/npcr/uscs/data/00_bias_correction.htm.
- 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 Census Geographic Areas Reference Manual, Chapter 6: Statistical Groupings of States and Counties for a list of states in each region.

Values	Frequency	Percentage
Northeast	5,520,861	20.7%
Midwest	6,008,531	22.5%
South	9,674,607	36.3%
West	5,454,071	20.5%

SEER*Stat Item Name: **USCS0116**

Source of Standard: NPCR

Source Item Name: Derived from NAACCR's *State at diagnosis (CoC)* and U.S. Cancer Statistics publication standards

Source Item Number: Derived from NAACCR's 80

Description

This variable indicates whether a central cancer registry met the U.S. Cancer Statistics publication standard for all cancer sites combined each year in 2001–2016.

Considerations for use

- This variable is used for analysis of combined 2001–2016 data in the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 database. Data from states that did not meet the U.S. Cancer Statistics publication standard in any given year were excluded from the database for that year. Please refer to Table 1 for the states available by year.
- If you are conducting a multiyear analysis and want to restrict the analysis to the states that met publication criteria for each of those years (for example, a trend analysis), we recommend using the predefined variables, *USCS1216* (includes diagnosis years 2012–2016), *USCS0716* (includes diagnosis years 2007–2016), or *USCS0116* (includes diagnosis years 2001–2016).
- 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.
- The U.S. Cancer Statistics publication standard is available at www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

Values	Frequency	Percentage
Does not meet USCS standard from 2001–2016	223,975	0.8%
Meets USCS standard from 2001–2016	26,434,095	99.2%

SEER*Stat Item Name: **USCS0716**

Source of Standard: NPCR

Source Item Name: Derived from NAACCR's *State at diagnosis (CoC)* and U.S. Cancer Statistics publication standard

Source Item Number: Derived from NAACCR's 80

Description

This variable indicates whether a central cancer registry met the U.S. Cancer Statistics publication standard for all cancer sites combined each year in 2007–2016. When using this variable, restrict the diagnosis years to 2007–2016. 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 2007–2016 data in the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 database. Data from states that did not meet the U.S. Cancer Statistics publication standard in any given year were excluded from the database for that year. Please refer to Table 1 for the states available by year.
- If you are conducting a multiyear analysis and want to restrict the analysis to the states that met publication criteria for each of those years (for example, a trend analysis), we recommend using the predefined variables, *USCS1216* (includes diagnosis years 2012–2016), *USCS0716* (includes diagnosis years 2007–2016), or *USCS0116* (includes diagnosis years 2001–2016).
- 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.
- The U.S. Cancer Statistics publication standard is available at www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

Note: This frequency table is restricted to cancers diagnosed during 2007–2016.

Values	Frequency	Percentage
Meets USCS standard from 2007–2016	15,727,417	100.0%

SEER*Stat Item Name: **USCS1216**

Source of Standard: NPCR

Source Item Name: Derived from NAACCR's *State at diagnosis (CoC)* and U.S. Cancer Statistics publication standard

Source Item Number: Derived from NAACCR's 80

Description

This variable indicates whether a central cancer registry met the U.S. Cancer Statistics publication standard for all cancer sites combined each year in 2012–2016. When using this variable, restrict the diagnosis years to 2012–2016. 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 2012–2016 data in the NPCR and SEER Incidence – U.S. Cancer Statistics 2001–2016 database. Data from states that did not meet the U.S. Cancer Statistics publication standard in any given year were excluded from the database for that year. Please refer to Table 1 for the states available by year.
- If you are conducting a multiyear analysis and want to restrict the analysis to the states that met publication criteria for each of those years (for example, a trend analysis), we recommend using the predefined variables, *USCS1216* (includes diagnosis years 2012–2016), *USCS0716* (includes diagnosis years 2007–2016), or *USCS0116* (includes diagnosis years 2001–2016).
- 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.
- The U.S. Cancer Statistics publication standard is available at www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

Note: This frequency table is restricted to cancers diagnosed during 2012–2016.

Values	Frequency	Percentage
Meets USCS standard from 2012–2016	9,016,192	100.0%

SEER*Stat Item Name: **Origin recode NHIA (Hispanic, Non-Hisp)**

Source of Standard: NAACCR
Source Item Name: *NHIA derived Hisp Origin*
Source Item Number: 191

Description

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]*.

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

Considerations for use

- This variable is available only in the 2001–2016 database.
- 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 state's 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 ethnicity cannot be displayed for Delaware, Kentucky, and Massachusetts.
  - Data for race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
- Blank values are allowed for states that chose not to include data for NHIA in this file.
- For more information, please see:
  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. Available at www.naaccr.org/wp-content/uploads/2016/11/NHIA_v2_2_1_09122011.pdf.

Values	Frequency	Percentage
Non-Spanish-Hispanic-Latino	24,848,668	93.2%
Spanish-Hispanic-Latino	1,806,023	6.8%
Invalid Value(s)	3,379	0.0%

SEER*Stat Item Name: **Primary Site – labeled**
Source of Standard: NAACCR
Source Item Name: Derived from *Primary site*
Source Item Number: 400

Description

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.
- For more information, please see SEER coding manuals http://seer.cancer.gov/tools/codingmanuals/historical.html.

Values	Frequency	Percentage
C00.0-C77.9 (specific site)	26,221,231	98.4%
C80.9 (Unknown primary site)	436,839	1.6%

SEER*Stat Item Name: **Histologic Type ICD-O-3**

Source of Standard: NAACCR
Source Item Name: *Histologic Type ICD-O-3*
Source Item Number: 522

Description

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 https://www.naaccr.org/wp-content/uploads/2016/11/2010-Implementation-Guidelines-and-Recommendations_Revised-June-2010.pdf.
- For more Information, please see:
  - 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
  - Surveillance, Epidemiology, and End Results Program. *Hematopoietic and Lymphoid Neoplasm Database*. Bethesda, MD: US Department of Health and Human Services, National Cancer Institute; 2016. https://seer.cancer.gov/seertools/hemelymph
  - Ruhl J, Adamo M, Dickie L. (January 2015). *Hematopoietic and Lymphoid Neoplasm Coding Manual*. National Cancer Institute, Bethesda, MD 20850-9765.
  - *International Classification of Diseases for Oncology*, Third Edition, First Revision. Geneva: World Health Organization, 2013: https://codes.iarc.fr/

Values	Frequency	Percentage
8000–9992	26,658,070	100.0%

SEER*Stat Item Name: **Grade**
Source of Standard: NAACCR
Source Item Name: *Grade*
Source Item Number: 440

Description

This variable indicates the grade or degree of differentiation of the primary tumor being reported. For lymphomas and leukemias, this field also is used to indicate T-, B-, Null-, or NK-cell origin.

Considerations for use

- 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 site-specific modules available at https://training.seer.cancer.gov/modules_site_spec.html and the most current FORDS manual (www.facs.org/cancer/coc/fordsmanual.html).

  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 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 made 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 in 2011 and later, cancer registries were required to report the World Health Organization (WHO) Grade Classification. Please see the variable description *CS Site-Specific Factor 1* for more information on this brain-specific grade classification.

Values	Frequency	Percentage
Well differentiated; Grade I	2,407,907	9.0%
Moderately differentiated; Grade II	6,563,626	24.6%
Poorly differentiated; Grade III	5,240,311	19.7%
Undifferentiated; anaplastic; Grade IV	869,278	3.3%
T-cell	93,443	0.4%
B-cell; pre-B; B-precursor	1,318,682	4.9%
Null cell; non T-non B	2,679	0.0%
NK cell; natural killer cell (1995+)	4,138	0.0%
Unknown	10,158,006	38.1%

SEER*Stat Item Name: **Diagnostic confirmation**

Source of Standard: NAACCR
Source Item Name: *Diagnostic confirmation*
Source Item Number: 490

Description

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, it is recommended to 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. Full incidence calculations must 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.
- For more information, please see:
  - FORDS at www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals
  - SEER coding manuals https://seer.cancer.gov/tools/codingmanuals/historical.html

Values	Frequency	Percentage
Microscopically confirmed (total)	25,026,429	93.9%
Positive histology	23,956,532	89.9%
Positive exfoliative cytology, no positive histology	823,864	3.1%
Positive histology AND immunophenotyping AND/OR positive genetic studies	216,656	0.8%
Positive microscopic confirm, method not specified	29,377	0.1%
Positive laboratory test/marker study	113,498	0.4%
Direct visualization without microscopic confirmation	34,908	0.1%
Radiography without microscopic confirm	959,276	3.6%
Clinical diagnosis only	157,492	0.6%
Unknown	366,467	1.4%

SEER*Stat Item Name: **ICD-O-3 Hist/behavior, labeled**

Source of Standard: SEER*Stat recode
Source Item Name: *ICD-O-3 Hist/behavior, labeled*
Source Item Number: Not applicable

Description

This variable includes 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 5-digit ICD-O-3 morphology code. The first 4 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 frequency table shown below. However, the morphology codes are not grouped in the database.
- For more information, please see
  - *International Classification of Diseases for Oncology.* Third Edition, First Revision. Geneva: World Health Organization, 2013.
  - SEER ICD-O-3 Coding Materials at https://seer.cancer.gov/icd-o-3.

ICD–O–3 Code	Label	Frequency	Percentage
800	Neoplasms, NOS	624,570	2.3%
801–804	Epithelial Neoplasms, NOS	1,599,549	6.0%
805–808	Squamous Cell Neoplasms	2,028,261	7.6%
809–811	Basal Cell Neoplasms	7,569	0.0%
812–813	Transitional Cell Papillomas and Carcinomas	1,155,586	4.3%
814–838	Adneomas and Adenocarcinomas	10,467,495	39.3%
839–842	Adnexal and Skin Appendage Neoplasms	30,443	0.1%
843	Mucoepidermoid Neoplasms	22,964	0.1%
844–849	Cystic, Mucinous and Serous Neoplasms	717,813	2.7%
850–854	Ductal and Lobular Neoplasms	3,987,270	15.0%
855	Acinar Cell Neoplasms	59,424	0.2%
856–857	Complex Epithelial Neoplasms	90,456	0.3%
858	Thymic Epithelial Neoplasms	12,200	0.0%
859–867	specialized Gonadal Neoplasms	7,592	0.0%
868–871	Paragangliomas and Glomus Tumors	4,362	0.0%
872–879	Nevi and Melanomas	1,785,808	6.7%
880	Soft Tissue Tumors and Sarcomas, NOS	55,901	0.2%
881–883	Fibromatous Neoplasms	57,510	0.2%
884	Myxomatous Neoplasms	1,330	0.0%
885–888	Lipomatous Neoplasms	39,512	0.1%
889–892	Myomatous Neoplasms	62,065	0.2%
893–899	Complex Mixed and Stromal Neoplasms	129,770	0.5%
900–903	Fibroepithelial Neoplasms	7,759	0.0%
904	Synovial-Like Neoplasms	9,433	0.0%
905	Mesothelial Neoplasms	50,318	0.2%
906–909	Germ Cell Neoplasms	141,589	0.5%
910	Trophoblastic Neoplasms	6,099	0.0%

ICD–O–3 Code	Label	Frequency	Percentage
911	Mesonephromas	284	0.0%
912–916	Blood Vessel Tumors	58,897	0.2%
917	Lymphatic Vessel Tumors	262	0.0%
918–924	Osseous and Chondromatous Neoplasms	32,264	0.1%
925	Giant Cell Tumors	1,377	0.0%
926	Miscellaneous Bone Tumors	8,373	0.0%
927–934	Odotogenic Tumors	908	0.0%
935–937	Miscellaneous Tumors	16,237	0.1%
938–948	Gliomas	320,171	1.2%
949–952	Neuroepitheliomatous Neoplasms	27,010	0.1%
953	Meningiomas	356,355	1.3%
954–957	Nerve Sheath Tumors	91,184	0.3%
958	Granular Cell Tumors and Alveolar Soft Part Sarcomas	1,226	0.0%
959–972	Hodgkin and Non-Hodgkin Lymphomas	1,111,171	4.2%
973	Plasma Cell Tumors	329,085	1.2%
974	Mast Cell Tumors	2,559	0.0%
975	Neoplasms of Histiocytes and Accessory Lymphoid Cells	6,769	0.0%
976	Immunoproliferative Disease	19,315	0.1%
980–994	Leukemias	711,748	2.7%
995–996	Chronic Myeloproliferative Disorders	160,057	0.6%
997	Other Hematologic Disorders	15,402	0.1%
998–999	Myelodysplastic Syndromes	224,768	0.8%

SEER*Stat Item Name: **Site recode ICD-O-3/WHO 2008**
Source of Standard: NAACCR
Source Item Name: Derived from *Primary site* and *Histologic code ICD-O-3*
Source Item Number: 400 (*Primary site*) and 522 (*Histologic code ICD-O-3*)

Description

This recode variable is defined by the SEER Program. The values of the site recode ICD-O-3/WHO 2008 variable 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 that are commonly used in the reporting of cancer incidence data.

Considerations for use

- This is the recommended variable for analyses by primary cancer site.
- More information, including the site recode number, is available at https://seer.cancer.gov/siterecode.

Values	Frequency	Percentage
All Sites (total)	26,658,070	100.0%
Oral Cavity and Pharynx	614,253	2.3%
Lip	37,035	0.1%
Tongue	180,087	0.7%
Salivary Gland	64,750	0.2%
Floor of Mouth	35,646	0.1%
Gum and Other Mouth	84,989	0.3%
Nasopharynx	28,159	0.1%
Tonsil	104,696	0.4%
Oropharynx	26,981	0.1%
Hypopharynx	37,308	0.1%
Other Oral Cavity and Pharynx	14,602	0.1%
Digestive System	4,467,972	16.8%
Esophagus	254,999	1.0%
Stomach	352,441	1.3%
Small Intestine	115,459	0.4%
Colon and Rectum	2,440,206	9.2%
Colon excluding Rectum	1,754,088	6.6%
Cecum	376,446	1.4%
Appendix	47,082	0.2%
Ascending Colon	336,906	1.3%
Hepatic Flexure	84,694	0.3%
Transverse Colon	159,686	0.6%
Splenic Flexure	53,606	0.2%
Descending Colon	103,799	0.4%
Sigmoid Colon	476,317	1.8%
Large Intestine, NOS	115,552	0.4%
Rectum and Rectosigmoid Junction	686,118	2.6%
Rectosigmoid Junction	177,271	0.7%
Rectum	508,847	1.9%
Anus, Anal Canal and Anorectum	104,750	0.4%
Liver and Intrahepatic Bile Duct	363,374	1.4%
Liver	322,002	1.2%

Values	Frequency	Percentage
Intrahepatic Bile Duct	41,372	0.2%
Gallbladder	60,099	0.2%
Other Biliary	89,068	0.3%
Pancreas	614,346	2.3%
Retroperitoneum	19,528	0.1%
Peritoneum, Omentum and Mesentery	30,105	0.1%
Other Digestive Organs	23,597	0.1%
Respiratory System	3,561,653	13.4%
Nose, Nasal Cavity and Middle Ear	36,266	0.1%
Larynx	212,803	0.8%
Lung and Bronchus	3,301,624	12.4%
Pleurae	1,506	0.0%
Trachea, Mediastinum and Other Respiratory Organs	9,454	0.0%
Bones and Joints	46,735	0.2%
Soft Tissue including Heart	164,033	0.6%
Skin excluding Basal and Squamous	1,821,731	6.8%
Melanoma of the Skin	1,737,375	6.5%
Other Non-Epithelial Skin	84,356	0.3%
Breast (female and male combined)	4,323,260	16.2%
Female Genital System	1,425,819	5.3%
Cervix Uteri	201,609	0.8%
Corpus and Uterus, NOS	727,821	2.7%
Corpus Uteri	706,637	2.7%
Uterus, NOS	21,184	0.1%
Ovary	339,612	1.3%
Vagina	23,417	0.1%
Vulva	103,844	0.4%
Other Female Genital Organs	29,516	0.1%
Male Genital System	3,458,989	13.0%
Prostate	3,292,126	12.3%
Testis	130,977	0.5%
Penis	29,770	0.1%
Other Male Genital Organs	6,116	0.0%
Urinary System	1,986,048	7.5%
Urinary Bladder	1,091,306	4.1%
Kidney and Renal Pelvis	831,402	3.1%
Ureter	43,558	0.2%
Other Urinary Organs	19,782	0.1%
Eye and Orbit	49,514	0.2%
Brain and Other Nervous System	824,399	3.1%
Brain	361,096	1.4%
Cranial Nerves Other Nervous System	463,303	1.7%
Endocrine System	814,649	3.1%
Thyroid	613,615	2.3%
Other Endocrine including Thymus	201,034	0.8%
Lymphoma	1,139,275	4.3%
Hodgkin Lymphoma	135,686	0.5%
Hodgkin – Nodal	131,952	0.5%
Hodgkin – Extranodal	3,734	0.0%

Values	Frequency	Percentage
Non-Hodgkin Lymphoma	1,003,589	3.8%
NHL – Nodal	681,141	2.6%
NHL – Extranodal	322,448	1.2%
Myeloma	326,275	1.2%
Leukemia	688,786	2.6%
Lymphocytic Leukemia	345,218	1.3%
Acute Lymphocytic Leukemia	75,546	0.3%
Chronic Lymphocytic Leukemia	247,479	0.9%
Other Lymphocytic Leukemia	22,193	0.1%
Myeloid and Monocytic Leukemia	307,718	1.2%
Acute Myeloid Leukemia	198,610	0.7%
Acute Monocytic Leukemia	11,988	0.0%
Chronic Myeloid Leukemia	88,157	0.3%
Other Myeloid/Monocytic Leukemia	8,963	0.0%
Other Leukemia	35,850	0.1%
Other Acute Leukemia	12,800	0.0%
Aleukemic, Subleukemic and NOS	23,050	0.1%
Mesothelioma	50,318	0.2%
Kaposi Sarcomae	20,347	0.1%
Miscellaneous	874,014	3.3%

SEER*Stat Item Name: **ICCC site recode ICD-O-3/WHO 2008**
Source of Standard: SEER
Source Item Name: Derived from NAACCR *Primary site*, *Histologic code ICD-O-3*, and *Behavior code ICD-O-3*
Source Item Number: NAACCR 400 (*Primary site*), 522 (*Histologic code ICD-O-3*), and 523 (*Behavior code ICD-O-3*)

Description

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

Considerations for use

- This recode is defined by the SEER Program, which was based on definitions presented in Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. *International Classification of Childhood Cancer*, Third Edition.
- For comparison of "*ICCC site rec extended ICD-O-3/WHO 2008* " and this variable see https://seer.cancer.gov/iccc/iccc-who2008.html.
- Additional information is available at https://seer.cancer.gov/iccc/iccc3.html and https://seer.cancer.gov/iccc/.

Note: This frequency table is restricted to individuals 19 years old or younger.

Values	Frequency	Percentage
I Leukemias, myeloproliferative & myelodysplastic diseases	61,779	23.3%
I(a) Lymphoid leukemias	44,055	16.6%
I(b) Acute myeloid leukemias	10,667	4.0%
I(c) Chronic myeloproliferative diseases	3,181	1.2%
I(d) Myelodysplastic syndrome and other myeloproliferative	1,996	0.8%
I(e) Unspecified and other specified leukemias	1,880	0.7%
II Lymphomas and reticuloendothelial neoplasms	35,590	13.4%
II(a) Hodgkin lymphomas	16,433	6.2%
II(b) Non-Hodgkin lymphomas (except Burkitt lymphoma)	11,917	4.5%
II(c) Burkitt lymphoma	3,222	1.2%
II(d) Miscellaneous lymphoreticular neoplasms	3,566	1.3%
II(e) Unspecified lymphomas	452	0.2%
III CNS and misc intracranial and intraspinal neoplasms	59,796	22.6%
III(a) Ependymomas and choroid plexus tumor	4,813	1.8%
III(b) Astrocytomas	22,070	8.3%
III(c) Intracranial and intraspinal embryonal tumors	8,049	3.0%
III(d) Other gliomas	7,472	2.8%
III(e) Other specified intracranial/intraspinal neoplasms	15,286	5.8%
III(f) Unspecified intracranial and intraspinal neoplasms	2,106	0.8%
IV Neuroblastoma and other peripheral nervous cell tumors	11,412	4.3%
IV(a) Neuroblastoma and ganglioneuroblastoma	11,093	4.2%
IV(b) Other peripheral nervous cell tumors	319	0.1%
V Retinoblastoma	4,241	1.6%
VI Renal tumors	9,154	3.5%
VI(a) Nephroblastoma and other nonepithelial renal tumors	8,242	3.1%
VI(b) Renal carcinomas	879	0.3%
VI(c) Unspecified malignant renal tumors	33	0.0%
VII Hepatic tumors	3,069	1.2%
VII(a) Hepatoblastoma	2,265	0.9%

Values	Frequency	Percentage
VII(b) Hepatic carcinomas	776	0.3%
VII(c) Unspecified malignant hepatic tumors	28	0.0%
VIII Malignant bone tumors	11,953	4.5%
VIII(a) Osteosarcomas	6,701	2.5%
VIII(b) Chondrosarcomas	455	0.2%
VIII(c) Ewing tumor and related sarcomas of bone	3,947	1.5%
VIII(d) Other specified malignant bone tumors	615	0.2%
VIII(e) Unspecified malignant bone tumors	235	0.1%
IX Soft tissue and other extraosseous sarcomas	16,062	6.1%
IX(a) Rhabdomyosarcomas	6,276	2.4%
IX(b) Fibrosarcomas, peripheral nerve & other fibrous	1,738	0.7%
IX(c) Kaposi sarcomae	61	0.0%
IX(d) Other specified soft tissue sarcomas	6,254	2.4%
IX(e) Unspecified soft tissue sarcomas	1,733	0.7%
X Germ cell & trophoblastic tumors & neoplasms of gonads	15,408	5.8%
X(a) Intracranial & intraspinal germ cell tumors	2,763	1.0%
X(b) Extracranial & extragonadal germ cell tumors	1,932	0.7%
X(c) Malignant gonadal germ cell tumors	9,697	3.7%
X(d) Gonadal carcinomas	574	0.2%
X(e) Other and unspecified malignant gonadal tumors	442	0.2%
XI Other malignant epithelial neoplasms and melanomas	25,449	9.6%
XI(a) Adrenocortical carcinomas	280	0.1%
XI(b) Thyroid carcinomas	10,973	4.1%
XI(c) Nasopharyngeal carcinomas	737	0.3%
XI(d) Malignant melanomas	6,694	2.5%
XI(e) Skin carcinomas	115	0.0%
XI(f) Other and unspecified carcinomas	6,650	2.5%
XII Other and unspecified malignant neoplasms	1,033	0.4%
XII(a) Other specified malignant tumors	547	0.2%
XII(b) Other unspecified malignant tumors	486	0.2%
Not classified by ICCC or *in situ*	9,723	3.7%

SEER*Stat Item Name: **ICCC site rec extended ICD-O-3/WHO 2008**
Source of Standard: SEER
Source Item Name: Derived from NAACCR *Primary site*, *Histologic code ICD-O-3*, and *Behavior code ICD-O-3*
Source Item Number: NAACCR 400 (*Primary site*), 522 (*Histologic code ICD-O-3*), and 523 (*Behavior code ICD-O-3)*

Description

This variable indicates the classification of childhood cancer, which is based on tumor morphology and primary site, and 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 (ICCC, 3rd Edition)* and *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (2008)*.

Considerations for use

- For comparison of *"ICCC site recode ICD-O-3/WHO 2008"* and this variable, please visit https://seer.cancer.gov/iccc/iccc-who2008.html.
- Additional information is available at http://seer.cancer.gov/iccc/iccc3_ext.html and https://seer.cancer.gov/iccc/.

Note: This frequency table is restricted to individuals 19 years old or younger.

Values	Frequency	Percentage
I Leukemias, myeloproliferative & myelodysplastic diseases	61,779	23.3%
I(a) Lymphoid leukemias	44,055	16.6%
I(a.1) Precursor cell leukemias	42,611	16.1%
I(a.2) Mature B-cell leukemias	1,100	0.4%
I(a.3) Mature T-cell and NK cell leukemias	150	0.1%
I(a.4) Lymphoid leukemia, NOS	194	0.1%
I(b) Acute myeloid leukemias	10,667	4.0%
I(c) Chronic myeloproliferative diseases	3,181	1.2%
I(d) Myelodysplastic syndrome and other myeloproliferative	1,996	0.8%
I(e) Unspecified and other specified leukemias	1,880	0.7%
II Lymphomas and reticuloendothelial neoplasms	35,590	13.4%
II(a) Hodgkin lymphomas	16,433	6.2%
II(b) Non-Hodgkin lymphomas (except Burkitt lymphoma)	11,917	4.5%
II(b.1) Precursor cell lymphomas	3,472	1.3%
II(b.2) Mature B-cell lymphomas except Burkitt lymphoma	4,783	1.8%
II(b.3) Mature T-cell and NK-cell lymphomas	3,005	1.1%
II(b.4) Non-Hodgkin lymphomas, NOS	657	0.2%
II(c) Burkitt lymphoma	3,222	1.2%
II(d) Miscellaneous lymphoreticular neoplasms	3,566	1.3%
II(e) Unspecified lymphomas	452	0.2%
III CNS and misc intracranial and intraspinal neoplasms	59,796	22.6%
III(a) Ependymomas and choroid plexus tumor	4,813	1.8%
III(a.1) Ependymomas	3,684	1.4%
III(a.2) Choroid plexus tumor	1,129	0.4%
III(b) Astrocytomas	22,070	8.3%
III(c) Intracranial and intraspinal embryonal tumors	8,049	3.0%
III(c.1) Medulloblastomas	5,203	2.0%
III(c.2) PNET	1,702	0.6%
III(c.3) Medulloepithelioma	81	0.0%

Values	Frequency	Percentage
III(c.4) Atypical teratoid/rhabdoid tumor	1,063	0.4%
III(d) Other gliomas	7,472	2.8%
III(d.1) Oligodendrogliomas	911	0.3%
III(d.2) Mixed and unspecified gliomas	6,404	2.4%
III(d.3) Neuroepithelial glial tumors of uncertain orig	157	0.1%
III(e) Other specified intracranial/intraspinal neoplasms	15,286	5.8%
III(e.1) Pituitary adenomas and carcinomas	6,452	2.4%
III(e.2) Tumors of sellar region (craniopharyngiomas)	2,191	0.8%
III(e.3) Pineal parenchymal tumors	615	0.2%
III(e.4) Neuronal and mixed neuronal-glial tumors	4,388	1.7%
III(e.5) Meningiomas	1,640	0.6%
III(f) Unspecified intracranial and intraspinal neoplasms	2,106	0.8%
IV Neuroblastoma and other peripheral nervous cell tumors	11,412	4.3%
IV(a) Neuroblastoma and ganglioneuroblastoma	11,093	4.2%
IV(b) Other peripheral nervous cell tumors	319	0.1%
V Retinoblastoma	4,241	1.6%
VI Renal tumors	^[1]	^[1]
VI(a) Nephroblastoma and other nonepithelial renal tumors	^[1]	^[1]
VI(a.1) Nephroblastoma	7,747	2.9%
VI(a.2) Rhabdoid renal tumor	211	0.1%
VI(a.3) Kidney sarcomas	272	0.1%
VI(a.4) pPNET of kidney	^[2]	^[2]
VI(b) Renal carcinomas	879	0.3%
VI(c) Unspecified malignant renal tumors	33	0.0%
VII Hepatic tumors	3,069	1.2%
VII(a) Hepatoblastoma	2,265	0.9%
VII(b) Hepatic carcinomas	776	0.3%
VII(c) Unspecified malignant hepatic tumors	28	0.0%
VIII Malignant bone tumors	11,953	4.5%
VIII(a) Osteosarcomas	6,701	2.5%
VIII(b) Chondrosarcomas	455	0.2%
VIII(c) Ewing tumor and related sarcomas of bone	3,947	1.5%
VIII(c.1) Ewing tumor and Askin tumor of bone	3,788	1.4%
VIII(c.2) pPNET of bone	159	0.1%
VIII(d) Other specified malignant bone tumors	615	0.2%
VIII(d.1) Malignant fibrous neoplasms of bone	57	0.0%
VIII(d.2) Malignant chordomas	291	0.1%
VIII(d.3) Odontogenic malignant tumors	73	0.0%
VIII(d.4) Miscellaneous malignant bone tumors	194	0.1%
VIII(e) Unspecified malignant bone tumors	235	0.1%
IX Soft tissue and other extraosseous sarcomas	16,062	6.1%
IX(a) Rhabdomyosarcomas	6,276	2.4%
IX(b) Fibrosarcomas, peripheral nerve & other fibrous	1,738	0.7%
IX(b.1) Fibroblastic and myofibroblastic tumors	908	0.3%
IX(b.2) Nerve sheath tumors	807	0.3%
IX(b.3) Other fibromatous neoplasms	23	0.0%
IX(c) Kaposi sarcomae	61	0.0%
IX(d) Other specified soft tissue sarcomas	6,254	2.4%
IX(d.1) Ewing tumor and Askin tumor of soft tissue	822	0.3%
IX(d.2) pPNET of soft tissue	375	0.1%
IX(d.3) Extrarenal rhabdoid tumor	342	0.1%

Values	Frequency	Percentage
IX(d.4) Liposarcomas	371	0.1%
IX(d.5) Fibrohistiocytic tumors	1,538	0.6%
IX(d.6) Leiomyosarcomas	258	0.1%
IX(d.7) Synovial sarcomas	1,382	0.5%
IX(d.8) Blood vessel tumors	239	0.1%
IX(d.9) Osseous & chondromatous neoplasms of soft tissue	127	0.0%
IX(d.10) Alveolar soft parts sarcoma	214	0.1%
IX(d.11) Miscellaneous soft tissue sarcomas	586	0.2%
IX(e) Unspecified soft tissue sarcomas	1,733	0.7%
X Germ cell & trophoblastic tumors & neoplasms of gonads	^1	^1
X(a) Intracranial & intraspinal germ cell tumors	2,763	1.0%
X(a.1) Intracranial & intraspinal germinomas	1,693	0.6%
X(a.2) Intracranial & intraspinal teratomas	711	0.3%
X(a.3) Intracranial & intraspinal embryonal carcinomas	39	0.0%
X(a.4) Intracranial & intraspinal yolk sac tumor	41	0.0%
X(a.5) Intracranial & intraspinal choriocarcinoma	25	0.0%
X(a.6) Intracranial & intraspinal tumors of mixed forms	254	0.1%
X(b) Extracranial & extragonadal germ cell tumors	1,932	0.7%
X(b.1) Germinomas: extracranial/extragonadal	205	0.1%
X(b.2) Malignant teratomas: extracranial/extragonadal	749	0.3%
X(b.3) Embryonal carcinomas: extracranial/extragonadal	22	0.0%
X(b.4) Yolk sac tumor: extracranial/extragonadal	439	0.2%
X(b.5) Choriocarcinomas: extracranial/extragonadal	219	0.1%
X(b.6) Other mixed germ cell: extracranial/extragonadal	298	0.1%
X(c) Malignant gonadal germ cell tumors	^1	^1
X(c.1) Malignant gonadal germinomas	2,034	0.8%
X(c.2) Malignant gonadal teratomas	1,681	0.6%
X(c.3) Gonadal embryonal carcinomas	948	0.4%
X(c.4) Gonadal yolk sac tumor	917	0.3%
X(c.5) Gonadal choriocarcinoma	99	0.0%
X(c.6) Malignant gonadal tumors of mixed forms	4,017	1.5%
X(c.7) Malignant gonadal gonadoblastoma	^2	^2
X(d) Gonadal carcinomas	574	0.2%
X(e) Other and unspecified malignant gonadal tumors	442	0.2%
XI Other malignant epithelial neoplasms and melanomas	25,449	9.6%
XI(a) Adrenocortical carcinomas	280	0.1%
XI(b) Thyroid carcinomas	10,973	4.1%
XI(c) Nasopharyngeal carcinomas	737	0.3%
XI(d) Malignant melanomas	6,694	2.5%
XI(e) Skin carcinomas	115	0.0%
XI(f) Other and unspecified carcinomas	6,650	2.5%
XI(f.1) Carcinomas of salivary glands	1,194	0.5%
XI(f.2) Carcinomas of colon and rectum	772	0.3%
XI(f.3) Carcinomas of appendix	1,444	0.5%
XI(f.4) Carcinomas of lung	602	0.2%
XI(f.5) Carcinomas of thymus	100	0.0%
XI(f.6) Carcinomas of breast	257	0.1%
XI(f.7) Carcinomas of cervix uteri	184	0.1%
XI(f.8) Carcinomas of bladder	387	0.1%
XI(f.9) Carcinomas of eye	37	0.0%
XI(f.10) Carcinomas of other specified sites	1,471	0.6%

Values	Frequency	Percentage
XI(f.11) Carcinomas of unspecified site	202	0.1%
XII Other and unspecified malignant neoplasms	^[1]	^[1]
XII(a) Other specified malignant tumors	^[1]	^[1]
XII(a.1) Gastrointestinal stromal tumor	88	0.0%
XII(a.2) Pancreatoblastoma	38	0.0%
XII(a.3) Pulmonary blastoma and pleuropulmonary blastoma	213	0.1%
XII(a.4) Other complex mixed and stromal neoplasms	42	0.0%
XII(a.5) Mesothelioma	36	0.0%
XII(a.6) Other specified malignant tumors	^[2]	^[2]
XII(b) Other unspecified malignant tumors	355	0.2%
Not classified by ICCC or *in situ*	8,557	4.1%

[1] Values are not reported due to the need for complementary cell suppression.

[2] Counts of fewer than 16 cases and the corresponding percentages are suppressed.

SEER*Stat Item Name: **AYA site recode ICD-O-3/WHO 2008**
Source of Standard: SEER
Source Item Name: Derived from NAACCR *Primary site*, *Histologic code ICD-O-3*, and *Behavior code ICD-O-3*
Source Item Number: NAACCR 400 (*Primary site*), 522 (*Histologic code ICD-O-3*), and 523 (*Behavior code ICD-O-3*)

Description

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 developed to better define the major cancer sites that affect adolescents or young adults between 15 and 29 years of age.

Considerations for use

This recode variable is defined by the SEER Program. It was based on definitions proposed by RD Barr and colleagues in Barr RD, Holowaty EJ, Birch JM. Classification Scheme for tumors diagnosed in adolescents and young adults. *Cancer* 2006;106(7):1425–30. For more information see https://seer.cancer.gov/ayarecode.

Note: This frequency table is restricted to individuals 15 to 29 years old.

Values	Frequency	Percentage
1 Leukemias	28,212	6.2%
1.1 Acute lymphoid leukemia	10,961	2.4%
1.2 Acute myeloid leukemia	10,650	2.3%
1.3 Chronic myeloid leukemia	4,432	1.0%
1.4 Other and unspecified leukemia	2,169	0.5%
2 Lymphomas	65,610	14.4%
2.1 Non-Hodgkin lymphoma	25,576	5.6%
2.2 Hodgkin lymphoma	40,034	8.8%
3 CNS and Oth Intracranial and Intraspinal Neo (all behav)	40,820	8.9%
3.1. Astrocytoma	13,618	3.0%
3.1.1 Specified low-grade astrocytic tumors	5,298	1.2%
3.1.2 Glioblastoma and anaplastic astrocytoma	5,158	1.1%
3.1.3 Astrocytoma, NOS	3,162	0.7%
3.2 Other glioma	7,170	1.6%
3.3 Ependymoma	2,640	0.6%
3.4. Medulloblastoma and other PNET	2,263	0.5%
3.4.1 Medulloblastoma	1,310	0.3%
3.4.2 Supratentorial PNET	953	0.2%
3.5 Other specified intracranial and intraspinal neoplasms	12,894	2.8%
3.6 Unspecified intracranial and intraspinal neoplasms	2,235	0.5%
3.6.1 Unspec malignant intracranial and intraspinal neo	405	0.1%
3.6.2 Unspec ben/border intracran. and intraspin neo	1,830	0.4%
4 Osseous & Chondromatous Neoplasms	11,623	2.5%
4.1 Osteosarcoma	5,020	1.1%
4.2 Chondrosarcoma	1,561	0.3%
4.3 Ewing tumor	3,937	0.9%
4.4 Other specified and unspecified bone tumors	1,105	0.2%
5 Soft Tissue Sarcomas	19,206	4.2%
5.1 Fibromatous neoplasms	4,742	1.0%
5.2 Rhabdomyosarcoma	2,155	0.5%
5.3 Other soft tissue sarcoma	12,309	2.7%

Values	Frequency	Percentage
5.3.1 Specified soft tissue sarcoma	9,530	2.1%
5.3.1.1 Specified (excluding Kaposi sarcoma)	7,614	1.7%
5.3.1.2 Kaposi sarcomae	1,916	0.4%
5.3.2 Unspecified soft tissue sarcoma	2,779	0.6%
6 Germ Cell and Trophoblastic Neoplasms	53,437	11.7%
6.1 Germ cell and trophoblastic neoplasms of gonads	48,820	10.7%
6.2 Germ cell and trophoblastic neo of nongonadal sites	4,617	1.0%
6.2.1 Intracranial (all behaviors)	1,763	0.4%
6.2.2 Other nongonadal	2,854	0.6%
7 Melanoma and Skin Carcinomas	41,100	9.0%
7.1 Melanoma	40,739	8.9%
7.2 Skin carcinomas	361	0.1%
8 Carcinomas	134,989	29.6%
8.1 Thyroid carcinoma	58,395	12.8%
8.2 Other carcinoma of head and neck	7,511	1.6%
8.2.1 Nasopharyngeal carcinoma	1,350	0.3%
8.2.2 Other sites in lip, oral cavity and pharynx	5,447	1.2%
8.2.3 Nasal cav, mid ear, sinus, larynx, ill-def head/neck	714	0.2%
8.3 Carcinoma of trachea, bronchus, and lung	3,150	0.7%
8.4 Carcinoma of breast	16,911	3.7%
8.5 Carcinoma of genitourinary tract	26,530	5.8%
8.5.1 Carcinoma of kidney	5,560	1.2%
8.5.2 Carcinoma of bladder	2,305	0.5%
8.5.3 Carcinoma of gonads	4,124	0.9%
8.5.4 Carcinoma of cervix and uterus	13,838	3.0%
8.5.5 Carc of oth and ill-defined sites	703	0.2%
8.6 Carcinoma of gastrointestinal tract	20,079	4.4%
8.6.1 Carcinoma of colon and rectum	13,570	3.0%
8.6.2 Carcinoma of stomach	1,908	0.4%
8.6.3 Carcinoma of liver and intrahepatic bile ducts	1,819	0.4%
8.6.4 Carcinoma of pancreas	1,421	0.3%
8.6.5 Carc oth and ill-def sites, gastrointestinal tract	1,361	0.3%
8.7 Carcinoma of other and ill-defined sites	2,413	0.5%
8.7.1 Adrenocortical carcinoma	385	0.1%
8.7.2 Carcinoma of other and ill-defined sites, NOS	2,028	0.4%
9 Miscellaneous specified neoplasms, NOS	10,578	2.3%
9.1 Other pediatric and embryonal tumors, NOS	1,055	0.2%
9.1.1 Wilms tumor	204	0.0%
9.1.2 Neuroblastoma	302	0.1%
9.1.3 Other pediatric and embryonal tumors, NOS	549	0.1%
9.2 Other specified and embryonal tumors, NOS	9,523	2.1%
9.2.1 Paraganglioma and glomus tumors	386	0.1%
9.2.2 Other specified gonadal tumors	778	0.2%
9.2.3 Myeloma, mast cell, misc lymphoreticular neo, NOS	1,720	0.4%
9.2.4 Other specified neoplasms, NOS	6,639	1.5%
10 Unspecified Malignant Neoplasms	2,447	0.5%
Unclassified and Non-Malignant	48,590	10.6%

SEER*Stat Item Name: **Lymphoma subtype recode/WHO 2008**
Source of Standard: SEER
Source Item Name: Derived from NAACCR *Primary site*, *Histologic code ICD-O-3*, and *Behavior code ICD-O-3*
Source Item Number: NAACCR 400 (*Primary site*), 522 (*Histologic code ICD-O-3*), and 523 (*Behavior code ICD-O-3*)

Description

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

- This recode variable is defined by the SEER Program. It was adapted from a proposed nested classification of lymphoid neoplasms in:
  Morton LM, et al. Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). *Blood* 2007;110:695–708.
- This variable is recommended to be used for analyses where the years of diagnosis are 2008 or later as it provides the most up-to-date definitions of lymphoma.
- More information is available at https://seer.cancer.gov/lymphomarecode.

Values	Frequency	Percentage
Lymphoid Neoplasm	1,839,973	6.9%
1 Hodgkin Lymphoma	135,686	0.5%
1(a) Classical Hodgkin lymphoma	128,222	0.5%
1(a)1 Lymphocyte-rich/mixed cell/lymphocyte depleted	21,631	0.1%
1(a)1.1 Lymphocyte-rich	4,767	0.0%
1(a)1.2 Mixed cellularity	15,231	0.1%
1(a)1.3 Lymphocyte-depleted	1,633	0.0%
1(a)2 Nodular sclerosis	72,819	0.3%
1(a)3 Classical Hodgkin lymphoma, NOS	33,772	0.1%
1(b) Nodular lymphocyte predominant Hodgkin lymphoma	7,464	0.0%
2 Non-Hodgkin lymphoma	1,651,383	6.2%
2(a) Non-Hodgkin lymphoma, B-cell	1,522,344	5.7%
2(a)1 Precursor Non-Hodgkin lymphoma, B-cell	59,484	0.2%
2(a)2 Mature Non-Hodgkin lymphoma, B-cell	1,382,189	5.2%
2(a)2.1 Chronic/Sm/Prolymphocytic/Mantle B-cell NHL	347,643	1.3%
2(a)2.1.1 Chronic/Small lymphocytic leuk/lymph	304,312	1.1%
2(a)2.1.2 Prolymphocytic leukemia, B-cell	1,378	0.0%
2(a)2.1.3 Mantle-cell lymphoma	41,953	0.2%
2(a)2.2 Lymphoplasmacytic lymphoma/Waldenstrom	32,128	0.1%
2(a)2.2.1 Lymphoplasmacytic lymphoma	13,416	0.1%
2(a)2.2.2 Waldenstrom macroglubulinemia	18,712	0.1%
2(a)2.3 Diffuse large B-cell lymphoma (DLBCL)	352,588	1.3%
2(a)2.3.1 DLBCL, NOS	348,854	1.3%
2(a)2.3.2 Intravascular large B-cell lymphoma	653	0.0%
2(a)2.3.3 Primary effusion lymphoma	462	0.0%

Values	Frequency	Percentage
2(a)2.3.4 Mediastinal large B-cell lymphoma	2,619	0.0%
2(a)2.4 Burkitt lymphoma/leukemia	19,662	0.1%
2(a)2.5 Marginal-zone lymphoma (MZL)	95,069	0.4%
2(a)2.5.1 Splenic MZL	8,805	0.0%
2(a)2.5.2 Extranodal MZL, MALT type	57,091	0.2%
2(a)2.5.3 Nodal MZL	29,173	0.1%
2(a)2.6 Follicular lymphoma	192,443	0.7%
2(a)2.7 Hairy-cell leukemia	14,950	0.1%
2(a)2.8 Plasma cell neoplasms	327,453	1.2%
2(a)2.8.1 Plasmacytoma	21,158	0.1%
2(a)2.8.2 Multiple myeloma/plasma-cell leuk	306,295	1.1%
2(a)2.9 Heavy chain disease	253	0.0%
2(a)3 Non-Hodgkin lymphoma, B-cell, NOS	80,671	0.3%
2(b) Non-Hodgkin lymphoma, T-cell	106,859	0.4%
2(b)1 Precursor Non-Hodgkin lymphoma, T-cell	4,789	0.0%
2(b)2 Mature Non-Hodgkin lymphoma, T-cell	101,505	0.4%
2(b)2.1 Mycosis fungoides/Sezary syndrome	22,331	0.1%
2(b)2.1.1 Mycosis fungoides	21,476	0.1%
2(b)2.1.2 Sezary syndrome	855	0.0%
2(b)2.2 Peripheral T-cell lymphoma	57,656	0.2%
2(b)2.2.1 Peripheral T-cell lymphoma, NOS	20,464	0.1%
2(b)2.2.2 Angioimmunoblastic T-cell lymphoma	6,395	0.0%
2(b)2.2.3 Subcutan panniculitis-like T-cell lymph	539	0.0%
2(b)2.2.4 Anaplastic lar cell lymph, T-/Null-cell	12,851	0.0%
2(b)2.2.5 Hepatosplenic T-cell lymphoma	472	0.0%
2(b)2.2.6 Enteropathy-type T-cell lymphoma	695	0.0%
2(b)2.2.7 Cutaneous T-cell lymphoma, NOS	11,812	0.0%
2(b)2.2.8 Prim cutaneous anaplastic lar cell lymph	4,428	0.0%
2(b)2.3 Adult T-cell leukemia/lymphoma	11,116	0.0%
2(b)2.4 NK/T-cell lymph, nasal-type/aggres NK leuk	3,142	0.0%
2(b)2.5 T-cell large granular lymphocytic leukemia	5,378	0.0%
2(b)2.6 Prolymphocytic leukemia, T-cell	1,882	0.0%
2(b)3 Non-Hodgkin lymphoma, NOS, T-cell	565	0.0%
2(c) Non-Hodgkin lymphoma, unknown lineage	22,180	0.1%
2(c)1 Precursor lymphoblastic leuk/lymph, unk lineage	8,352	0.0%
2(c)2 Prolymphocytic leukemia, unknown lineage	655	0.0%
2(c)3 Non-Hodgkin lymphoma, NOS, unknown lineage	13,173	0.0%
3 Composite Hodgkin lymphoma and NHL	3,527	0.0%
4 Lymphoid neoplasm, NOS	49,377	0.2%
Unclassified	24,818,097	93.1%

SEER*Stat Item Name: **Behavior Recode for analysis derived/WHO2008**
Source of Standard: NAACCR
Source Item Name: *Behavior code ICD-O-3*
Source Item Number: 523

Description

This variable is the behavior code from *The International Classification of Diseases for Oncology*, Third Edition (ICD-O-3) for the primary tumor being reported. With the updates and implementation of ICD-O, some histologies changed from malignant behavior to benign or borderline malignancy and vice versa. This Behavior Recode variable takes those changes into consideration.

"Malignant" indicates a histology whose behavior did not change. "Only malignant in ICD-O-3" indicates histologies whose behavior changed to malignant in edition ICD-O-3. Additional changes to histologic behavior were made effective with cases diagnosed in 2010 and later; new histologies were added, while the behavior of others changed (primarily affecting hematopoetic malignancies). "Only malignant 2010+" limits the analysis to those histologies.

Considerations for use

- This database includes cases with invasive (malignant) and *in situ* behavior, benign/borderline behavior for brain/CNS cases diagnosed in 2004 and forward are also included. Caution should be used to select the correct behavior for analyses. Malignant behavior (specifically the "Malignant" category) is the default selection for cases in this database in SEER*Stat. If necessary for the analysis, "Only malignant in ICD-O-3" or "Only malignant 2010+" may be selected to further restrict case selection. If an analysis requires cases with *in situ* behavior, the "Malignant Only" selection should be unchecked on the "Selection" tab.
- *Behavior code ICD-O-3* is required for cancer cases diagnosed on or after January 1, 2001. Statistics generated from the database using this behavior code variable are comparable to the methodology used to generate the U.S. Cancer Statistics official federal cancer statistics.
- For more information, please see SEER coding manual at http://seer.cancer.gov/icd-o-3.

Values	Frequency	Percentage
Benign	598,390	2.2%
Borderline malignancy	58,124	0.2%
In situ	1,843,564	6.9%
Malignant	23,751,361	89.1%
Only malignant in ICD-O-3	382,410	1.4%
Only malignant 2010+	24,221	0.1%

SEER*Stat Item Name: **Merged Summary Stage 2000**
Source of Standard: NPCR
Source Item Name: Combined from *Derived SS2000* and *SEER Summary Stage 2000*
Source Item Number: Derived from NAACCR 3020 (*Derived SS2000*) and 759 (*SEER Summary Stage 2000*)

Description

This is a merged stage variable created using two variables: *SEER Summary Stage 2000*, which records stage from diagnosis years 2001–2003, *Derived SS2000*, which records stage from diagnostic years 2004–2015, and *SEER Summary Stage 2000* for diagnostic year 2016. This stage variable can be used for diagnosis years 2001–2016.

Considerations for use

- The coding logic for this merged variable is:
  - If a case was diagnosed between 2001 and 2003, the *Summary Stage 2000* variable value was used.
  - If a case was diagnosed between 2004 and 2015, then the Derived Summary Stage 2000 (*Derived SS2000*) variable was used.
  - If a case was diagnosed in 2016, then the *SEER Summary Stage 2000* variable was used.
  - If the Derived Summary Stage 2000 variable was blank and a valid value was available for the *Summary Stage 2000* variable, that value was used to populate the merged variable. For example, if the case was diagnosed in 2013 and *Derived Summary Stage 2000* was blank, but *Summary Stage 2000* had a value of local, then the merged variable was coded as local stage. Otherwise, the merged variable would be left blank.
- For more information about *SEER Summary Stage 2000* and *Derived SS2000* variables, see https://cancerstaging.org/cstage/Pages/default.aspx.

Values	Frequency	Percentage
In situ	2,294,039	8.6%
Localized only	10,887,616	40.8%
Regional, direct extension only	1,756,768	6.6%
Regional, regional lymph nodes only	1,882,787	7.1%
Regional, direct extension and regional lymph nodes	1,101,571	4.1%
Regional, NOS	271,504	1.0%
Distant site(s)/node(s) involved	5,836,927	21.9%
Not applicable	637,229	2.4%
Unknown/unstaged/unspecified	1,989,198	7.5%
Blanks	431	0.0%

SEER*Stat Item Name: **RX Summ – Surg Prim Site**
Source of Standard: SEER / CoC
Source Item Name: *RX Summ—Surg Prim Site*
Source Item Number: 1290

Description

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** starting with **diagnosis year 2003**.
  - For breast surgery codes, refer to the *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
- In **addition to the site-specific codes, refer to the most recent version of *FORDS* and SEER Program Code manual for additional instructions:**
  - *FORDS* manual - https://www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals
  - SEER Program Code manual - https://seer.cancer.gov/tools/codingmanuals

Note: This frequency table is restricted to female breast cases and diagnosis years ≥2003.

Values	Frequency	Percentage
00-99	3,803,797	100.0%

SEER*Stat Item Name: **CS Site-Specific Factor 1**
Source of Standard: AJCC
Source Item Name: *CS Site-Specific Factor 1*
Source Item Number: 2880

Description

The information recorded in *CS Site-Specific Factor 1* (SSF1) differs for each anatomic site and there is site-specific codes and coding structures for each anatomic site. In the U.S. Cancer Statistics Incidence Analytic database, SSF1 records information represents:

- **Female breast site: Estrogen Receptor (ER) Assay.**
- **Brain site: World Health Organization (WHO) Grade Classification**.

Considerations for use

- Data for this variable are available for:
  - **Female breast** starting with **diagnosis year 2004**.
  - **Brain** starting with **diagnosis year 2011**.
- For the site-specific codes, please refer to the Collaborative Stage Data Collection System
  - **Female breast:** Breast Estrogen Receptor Assay, available at https://web2.facs.org/cstage0205/breast/Breast_jag.html.
  - **Brain:** World Health Organization (WHO) Grade Classification, available at https://web2.facs.org/cstage0205/brain/Brain_jpo.html.

Note: This frequency table is restricted to female breast cases and diagnosis years ≥2004.

Values	Frequency	Percentage
00-99	3,546,828	99.5%
Blank(s)	18,298	0.5%

Note: This frequency table is restricted to brain cases and diagnosis years ≥2011.

Values	Frequency	Percentage
00-99	147,517	99.9%
Blank(s)	214	0.1%

SEER*Stat Item Name: **CS Site-Specific Factor 2**
Source of Standard: AJCC
Source Item Name: *CS Site-Specific Factor 2*
Source Item Number: 2890

Description

The information recorded in CS Site-Specific Factor 2 (SSF2) differs for each anatomic site and there is site-specific codes and coding structures for each anatomic site.

In the U.S. Cancer Statistics Incidence Analytic database, SSF2 records information for **female breast**, specifically, **Progesterone Receptor (PR) Assay**.

Considerations for use

- Data for this variable are available for **female breast** starting with **diagnosis year 2004**.
- Please refer to the Collaborative Stage Data Collection System for the specific codes for CS SSF2, Breast Progesterone Receptor Assay, available at http://web2.facs.org/cstage0205/breast/Breast_kac.html.

Note: This frequency table is restricted to female breast cases and diagnosis years ≥2004.

Values	Frequency	Percentage
00-99	3,546,824	99.5%
Blank(s)	18,302	0.5%

SEER*Stat Item Name: **CS Site-Specific Factor 15**
Source of Standard: AJCC
Source Item Name: *CS Site-Specific Factor 15*
Source Item Number: 2869

Description

The information recorded in CS Site-Specific Factor 15 (SSF15) differs for each anatomic site and there is site-specific codes and coding structures for each anatomic site.

In the U.S. Cancer Statistics Incidence Analytic database, SSF2 records information for **female breast**, specifically, **Human Epidermal Growth Factor Receptor 2 (HER2): Summary Result of Testing**.

Considerations for use

- Data for this variable are available for **female breast** starting with **diagnosis year 2010**.
- Please refer to the Collaborative Stage Data Collection System for the specific codes for CS SSF3, Breast HER2, available at http://web2.facs.org/cstage0205/breast/Breast_sbg.html.

Note: This frequency table is restricted to female breast cases and diagnosis years ≥2010.

Values	Frequency	Percentage
00-99	2,017,471	99.8%
Blank(s)	4,491	0.2%

SEER*Stat Item Name: **Laterality**

Source of Standard: NAACCR
Source Item Name: *Laterality at Diagnosis (SEER)*
Source Item Number: 410

Description

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.

Considerations for use

For more information, please see:

- FORDS at www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals
- SEER coding manuals http://seer.cancer.gov/tools/codingmanuals/historical.html

Values	Frequency	Percentage
Not a paired site	14,620,659	54.8%
Right - origin of primary	5,831,090	21.9%
Left - origin of primary	5,402,396	20.3%
Only one side - side unspecified	53,393	0.2%
Bilateral, single primary	204,500	0.8%
Paired site: midline tumor	61,106	0.2%
Paired site, but no information concerning laterality	484,926	1.8%

SEER*Stat Item Name: **Sequence Number – Central**
Source of Standard: NAACCR
Source Item Name: *Sequence Number – Central Revised*
Source Item Number: 380

Description

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. Please see the SEER list at https://seer.cancer.gov/tools/casefinding/. 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* may also affect the sequence number. For more information, please see https://seer.cancer.gov/tools/mphrules/mphrules_instructions.pdf.
- For more information, please see the SEER coding manual at https://seer.cancer.gov/tools/codingmanuals/historical.html.

Values	Frequency	Percentage
One primary only	19,076,875	^[1]
1st of 2 or more primaries	2,317,766	^[1]
2nd of 2 or more primaries	3,805,664	^[1]
3rd of 3 or more primaries	649,103	^[1]
4th of 4 or more primaries	115,635	^[1]
5th of 5 or more primaries	24,245	^[1]
6th or more primaries[2]	11,745	^[1]
Only one state registry-defined neoplasm	627,501	^[1]
1st of 2 or more state registry-defined neoplasms	12,857	^[1]
2nd of 2 or more state registry-defined neoplasms	14,528	^[1]
3rd of 3 or more state registry-defined neoplasms	1,024	^[1]
4th of 4 or more state registry-defined neoplasms	264	^[1]
5th of 5 or more state registry-defined neoplasms	111	^[1]
6th or more state registry-defined neoplasms[1]	71	^[1]

Values	Frequency	Percentage
Unknown sequence number - federally required *in situ* or malignant tumors	118	^[1]
Carcinoma *in situ* of the Cervix diagnosed 1/1/1996 or later	^[3]	^[3]
Unknown sequence number - state registry-defined neoplasms	385	^[1]

[1] Values are not reported due to the need for complementary cell suppression.

[2] Subsequent primaries (7 or higher) were collapsed into this category.

[3] Counts of fewer than 16 cases and the corresponding percentages are suppressed.

SEER*Stat Item Name: **NHIA Derived Hisp Origin**

Source of Standard: NAACCR
Source Item Name: *NHIA Derived Hispanic Origin*
Source Item Number: 191

Description

The NAACCR Hispanic Identification Algorithm uses a combination of data items to directly or indirectly classify cases as Hispanic for analytic purposes. Cases are classified based on individual's birth place (Non-Hispanic, Mexican, Puerto Rican, Cuban).

Considerations for use

- **This variable includes only count data**; rates cannot be calculated using this variable, as no population data are associated with the variable. Use the variable *Origin recode NHIA (Hispanic/Non-Hispanic)* to calculate age-adjusted rates by ethnicity.
- Blank values are allowed for states that chose not to include data for NHIA in this file.
- The following states have state-level race or ethnicity data presentation restrictions:
  - Data for Hispanic ethnicity cannot be displayed for Delaware, Kentucky, and Massachusetts.
  - Data for race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.

Values	Frequency	Percentage
Non-Spanish-Hispanic-Latino	24,848,668	93.2%
Mexican	354,357	1.3%
Puerto Rican	106,114	0.4%
Cuban	75,516	0.3%
South or Central American excluding Brazil	143,269	0.5%
Other specified Spanish/Hispanic origin including Europe	50,792	0.2%
Spanish/Hispanic/Latino, NOS	791,309	3.0%
NHIA surname match only	250,464	0.9%
Dominican Republic	34,202	0.1%
Invalid Value(s)	3,379	0.0%

SEER*Stat Item Name: **Year of Birth**
Source of Standard: SEER / CoC
Source Item Name: Date of Birth
Source Item Number: 240

Description

Year of birth of the patient.

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 should be 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	Frequency	Percentage
1890	^1	^1
1891	^1	^1
1892	^1	^1
1893	^1	^1
1894	16	^2
1895	28	^2
1896	46	^2
1897	70	^2
1898	146	^2
1899	226	^2
1900	556	^2
1901	853	^2
1902	1,269	^2
1903	1,893	^2
1904	2,953	^2
1905	4,552	^2
1906	6,541	^2
1907	10,006	^2
1908	14,377	^2
1909	19,442	^2
1910	27,082	^2
1911	35,155	^2
1912	48,466	^2
1913	61,332	^2
1914	79,367	^2
1915	95,635	^2
1916	116,204	^2
1917	141,529	^2
1918	172,494	^2

Values	Frequency	Percentage
1919	196,375	^2
1920	245,100	^2
1921	288,914	^2
1922	315,529	^2
1923	352,401	^2
1924	395,270	^2
1925	422,728	^2
1926	453,117	^2
1927	492,272	^2
1928	509,894	^2
1929	523,771	^2
1930	556,706	^2
1931	554,248	^2
1932	568,788	^2
1933	557,021	^2
1934	589,707	^2
1935	604,560	^2
1936	610,299	^2
1937	624,158	^2
1938	643,246	^2
1939	633,772	^2
1940	643,593	^2
1941	660,902	^2
1942	712,504	^2
1943	711,447	^2
1944	654,266	^2
1945	618,508	^2
1946	702,645	^2
1947	746,443	^2
1948	673,580	^2
1949	636,171	^2
1950	595,882	^2
1951	583,322	^2
1952	564,142	^2
1953	535,144	^2
1954	519,330	^2
1955	489,529	^2
1956	470,787	^2
1957	448,153	^2
1958	412,042	^2
1959	385,329	^2
1960	358,409	^2
1961	329,235	^2

Values	Frequency	Percentage
1962	299,081	^2
1963	272,951	^2
1964	246,124	^2
1965	213,686	^2
1966	189,292	^2
1967	168,602	^2
1968	155,757	^2
1969	146,239	^2
1970	138,207	^2
1971	121,924	^2
1972	105,055	^2
1973	91,880	^2
1974	85,287	^2
1975	77,996	^2
1976	71,662	^2
1977	67,302	^2
1978	61,875	^2
1979	59,522	^2
1980	55,265	^2
1981	51,155	^2
1982	47,095	^2
1983	42,175	^2
1984	38,898	^2
1985	35,428	^2
1986	32,142	^2
1987	29,218	^2
1988	27,248	^2
1989	24,607	^2
1990	22,649	^2
1991	20,342	^2
1992	18,150	^2
1993	16,305	^2
1994	14,834	^2
1995	13,384	^2
1996	12,535	^2
1997	12,007	^2
1998	12,441	^2
1999	11,881	^2
2000	11,895	^2
2001	11,479	^2
2002	10,675	^2
2003	10,078	^2
2004	9,479	^2

Values	Frequency	Percentage
2005	8,736	^[2]
2006	8,510	^[2]
2007	8,025	^[2]
2008	7,566	^[2]
2009	6,701	^[2]
2010	6,109	^[2]
2011	5,318	^[2]
2012	4,624	^[2]
2013	3,962	^[2]
2014	2,780	^[2]
2015	1,736	^[2]
2016	694	^[2]
Blank(s)	^[1]	^[1]

[1] Values are not reported due to the need for complementary cell suppression.

[2] Counts of fewer than 16 cases and the corresponding percentages are suppressed.

SEER*Stat Item Name: **Month of Diagnosis**

Source of Standard: NAACCR
Source Item Name: Derived from *Date of initial diagnosis (CoC)*
Source Item Number: 390

Description

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	Frequency	Percentage
January	2,315,500	8.7%
February	2,055,762	7.7%
March	2,251,844	8.4%
April	2,212,977	8.3%
May	2,221,624	8.3%
June	2,278,404	8.5%
July	2,166,587	8.1%
August	2,228,581	8.4%
September	2,125,663	8.0%
October	2,256,636	8.5%
November	2,070,801	7.8%
December	2,048,619	7.7%
Blank(s)	425,072	1.6%

SEER*Stat Item Name: **Type of Reporting Source**

Source of Standard: NAACCR
Source Item Name: *Type of reporting source*
Source Item Number: 500

Description

This variable codes the source documents used to abstract the majority of information on the tumor being reported. This may not be the source of original casefinding (for example, if a case is identified through a pathology laboratory report review and all source documents used to abstract the case are from the physician's office, code 4 is assigned).

Considerations for use

- For cancers diagnosed prior to 2006, only the following categories were available for *Type of Reporting Source*:

Code	Definition
1	Hospital inpatient/outpatient or clinic
3	Laboratory only (hospital or private)
4	Physician's office/private medical practitioner (local medical doctor)
5	Nursing/convalescent home/hospice

  Since there was no code specific to radiation treatment facilities or ambulatory surgery centers, they were coded as either 1 or 4.

- For cancers diagnosed in 2006 and later, the following codes were added to allow identification of cases reported by radiation treatment centers, medical oncology clinics, and other hospital outpatient units/surgery centers.

Code	Definition
2	Radiation treatment centers, medical oncology clinics
8	Other hospital outpatient units/surgery centers

- Cases were coded in the following priority order: 1, 2, 8, 4, 3, 5. This reflects the addition of codes 2 and 8 and prioritizes laboratory reports over nursing home reports. The source facilities included in the previous code 1 (hospital inpatient and outpatient) are split between codes 1, 2, and 8.

Values	Frequency	Percentage
Hospital inpatient/outpatient or clinic	23,188,579	87.0%
Radiation treatment or medical oncology center (2006+)	546,632	2.1%
Laboratory only (hospital or private)	717,446	2.7%
Physician's office/private medical practitioner (LMD)	1,163,625	4.4%
Nursing/convalescent home/hospice	35,292	0.1%
Other hospital outpatient unit or surgery center (2006+)	1,006,496	3.8%

SEER*Stat Item Name: **Alcohol-Related Cancers**
Source of Standard: NPCR
Source Item Name: Derived from NAACCR's *Primary Site*, *Histologic Type ICD-O-3*, *Sex*
Source Item Number: Derived from NAACCR's 400 (*Primary Site*), 522 (*Histologic Type ICD-O-3*), 220 (*Sex*)

Description

Predefined variable created using ICD-O-3 site, histology and sex to define alcohol-related cancers[3,4].

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[5,6,7].
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, also known as the *attributable fraction*[8]. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information on risk-factor related cancers, please see the publication listed below and online for the variable definitions:
  - 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. DOI: http://dx.doi.org/10.15585/mmwr.mm6603a1.
  - *Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors* documentation available at https://www.cdc.gov/cancer/public-use.

  Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers with the exception of bladder cancers (includes *in situ* and invasive cancers).

Values	Frequency	Percentage
Lip, oral cavity, pharynx	614,253	7.6%
Esophagus	254,999	3.1%
Colon and rectum	2,440,206	30.0%
Liver	322,002	4.0%
Larynx	212,803	2.6%
Female breast cancer	4,289,938	52.7%

[3] 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. Available at http://monographs.iarc.fr/ENG/Monographs/vol96/.
[4] 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. Available at http://monographs.iarc.fr/ENG/Monographs/vol100E/.
[5] Henley S.J., Singh S.D., King J, et al. Invasive cancer incidence and survival—United States, 2013. *MMWR* 2017.
[6] Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *Journal of the National Cancer Institute* 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? *Preventing Chronic Disease* 2007;4(1):A14.

SEER*Stat Item Name: **HPV-Related Cancers**

Source of Standard: NPCR

Source Item Name: Derived from NAACCR's *Primary Site*, *Histologic Type ICD-O-3*, *Sex*, *Diagnostic Confirmation*

Source Item Number: Derived from NAACCR's 400 (*Primary Site*), 522 (*Histologic Type ICD-O-3*), 220 (*Sex*), 490 (*Diagnostic Confirmation*)

Description

Predefined variable created using ICD-O-3 site, histology and sex to define Human Papillomavirus (HPV)-related cancers[9,10,11,12,13].

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[14,15,16].
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, also known as the *attributable fraction*[17]. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information on risk-factor related cancers, please see the publication listed below and online for the variable definitions:
  - 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. DOI: http://dx.doi.org/10.15585/mmwr.mm6603a1.
  - *Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors* documentation available at https://www.cdc.gov/cancer/public-use.

  Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers with the exception of bladder cancers (includes *in situ* and invasive cancers).

Values	Frequency	Percentage
Oropharyngeal squamous cell carcinoma	244,084	36.9%
Anal and rectal squamous cell carcinoma	100,163	15.1%
Vulvar squamous cell carcinoma	80,140	12.1%
Vaginal squamous cell carcinoma	15,447	2.3%

[9] Watson M, Saraiya M, Ahmed F, Cardinez CJ, Reichman ME, Weir HK, Richards TB. 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. Available at www.ncbi.nlm.nih.gov/pubmed/18980203.

[10] Saraiya M, Unger ER, Thompson TD, Lynch CF, Hernandez BY, Lyu CW,Steinau M, Watson M, Wilkinson EJ, Hopenhayn C, Copeland G, Cozen W, Peters ES, Huang Y, Saber MS, Altekruse S, Goodman MT; HPV Typing of Cancers Workgroup. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. *Journal of the National Cancer Institute* 2015;107(6):djv086. Available at www.ncbi.nlm.nih.gov/pubmed/25925419.

[11] 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. Available at http://monographs.iarc.fr/ENG/Monographs/vol90/.

[12] Viens LJ, Henley SJ, Watson M, Markowitz LE, Thomas CC, Thompson TD, Razzaghi H, Saraiya M, Centers for Disease Control and Prevention (CDC). Human papillomavirus–associated cancers—United States, 2008–2012. *MMWR* 2016;65(26):661–666. Available at www.cdc.gov/mmwr/volumes/65/wr/mm6526a1.htm.

[13] Centers for Disease Control and Prevention. How Many Cancers Are Linked with HPV Each Year? Atlanta, GA: U.S. Department of Health and Human Services. Available at www.cdc.gov/cancer/hpv/statistics/cases.htm.

[14] Henley S.J., Singh S.D., King J, et al. Invasive cancer incidence and survival—United States, 2013. *MMWR* 2017.

[15] Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *Journal of the National Cancer Institute* 2011;103(24):1827–1839.

[16] 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.

[17] Levine B. What does the population attributable fraction mean? *Preventing Chronic Disease* 2007;4(1):A14.

Penile squamous cell carcinoma	27,811	4.2%
Cervical carcinoma	193,629	29.3%

SEER*Stat Item Name: **Obesity-Related Cancers**
Source of Standard: NPCR
Source Item Name: Derived from NAACCR's *Primary Site*, *Histologic Type ICD-O-3*, *Sex*, *Diagnostic Confirmation*, Age at diagnosis
Source Item Number: Derived from NAACCR's 400 (*Primary Site*), 522 (*Histologic Type ICD-O-3*), 220 (*Sex*), 490 (*Diagnostic Confirmation*), 230 (*Age at diagnosis*)

Description

Predefined variable created using ICD-O-3 site, histology and sex to define obesity-related cancers[18,19,20].

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[21,22,23].
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, also known as the *attributable fraction*[24]. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information on risk-factor related cancers, please see the publication listed below and online for the variable definitions:
  - 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. DOI: http://dx.doi.org/10.15585/mmwr.mm6603a1.
  - *Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors* documentation available at https://www.cdc.gov/cancer/public-use.

  Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers with the exception of bladder cancers (includes *in situ* and invasive cancers).

Values	Frequency	Percentage
Esophageal adenocarcinoma	155,859	1.5%
Gastric cardia	107,969	1.1%
Colon & rectum	2,440,206	23.9%
Liver	322,002	3.2%
Gallbladder	60,099	0.6%
Pancreas	614,346	6.0%
Kidney	770,960	7.6%
Meningioma	351,420	3.4%
Thyroid	613,615	6.0%
Multiple myeloma	304,337	3.0%

[18] Eheman C, Henley SJ, Ballard-Barbash R, Jacobs EJ, Schymura MJ, Noone AM, Pan L, Anderson RN, Fulton JE, Kohler BA, Jemal A, Ward E, Plescia M, Ries LA, Edwards BK. 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. Available at www.ncbi.nlm.nih.gov/pmc/articles/PMC4586174/.
[19] 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. Available at http://preventcancer.aicr.org/site/PageServer?pagename=research_science_expert_report.
[20] 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.
[21] Henley S.J., Singh S.D., King J, et al. Invasive cancer incidence and survival—United States, 2013. *MMWR* 2017.
[22] Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *Journal of the National Cancer Institute* 2011;103(24):1827–1839.
[23] 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.
[24] Levine B. What does the population attributable fraction mean? *Preventing Chronic Disease* 2007;4(1):A14.

Postmenopausal female breast	3,391,073	33.2%
Corpus and uterus, NOS (not otherwise specified)	727,821	7.1%
Ovary	339,612	3.3%

SEER*Stat Item Name: **Physical Inactivity-Related Cancers**
Source of Standard: NPCR
Source Item Name: Derived from NAACCR's *Primary Site*, *Histologic Type ICD-O-3*, *Sex*
Source Item Number: Derived from NAACCR's 400 (*Primary Site*), 522 (*Histologic Type ICD-O-3*), 220 (*Sex*)

Description

Predefined variable created using ICD-O-3 site, histology and sex to define physical inactivity-related cancers[25,26].

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[27,28,29].
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, also known as the *attributable fraction*[30]. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information on risk-factor related cancers, please see the publication listed below and online for the variable definitions:
  - 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. DOI: http://dx.doi.org/10.15585/mmwr.mm6603a1.
  - *Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors* documentation available at https://www.cdc.gov/cancer/public-use.

  Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers with the exception of bladder cancers (includes *in situ* and invasive cancers).

Values	Frequency	Percentage
Colon	1,754,088	29.9%
Postmenopausal female breast	3,391,073	57.7%
Corpus and uterus, NOS (not otherwise specified)	727,821	12.4%

[25] Eheman C, Henley SJ, Ballard-Barbash R, Jacobs EJ, Schymura MJ, Noone AM, Pan L, Anderson RN, Fulton JE, Kohler BA, Jemal A, Ward E, Plescia M, Ries LA, Edwards BK. 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. Available at www.ncbi.nlm.nih.gov/pmc/articles/PMC4586174/.
[26] 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. Available at http://preventcancer.aicr.org/site/PageServer?pagename=research_science_expert_report.
[27] Henley S.J., Singh S.D., King J, et al. Invasive cancer incidence and survival—United States, 2013. *MMWR* 2017.
[28] Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *Journal of the National Cancer Institute* 2011;103(24):1827–1839.
[29] 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.
[30] Levine B. What does the population attributable fraction mean? *Preventing Chronic Disease* 2007;4(1):A14.

SEER*Stat Item Name: **Tobacco-Related Cancers**
Source of Standard: NPCR
Source Item Name: Derived from NAACCR's *Primary Site*, *Histologic Type ICD-O-3*, *Sex*
Source Item Number: Derived from NAACCR's 400 (*Primary Site*), 522 (*Histologic Type ICD-O-3*), 220 (*Sex*)

Description

Predefined variable created using ICD-O-3 site, histology and sex to define tobacco-related cancers[31].

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[32,33,34].
- However, other sources of information can be used to obtain the proportion of cancers probably caused by the risk factor, also known as the *attributable fraction*[35]. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information on risk-factor related cancers, please see the publication listed below and online for the variable definitions:
  - 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. DOI: http://dx.doi.org/10.15585/mmwr.mm6603a1.
  - *Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors* documentation available at https://www.cdc.gov/cancer/public-use.

  Please note that in official federal cancer statistics publications, CDC reports malignant (invasive) cancers with the exception of bladder cancers (includes *in situ* and invasive cancers).

Values	Frequency	Percentage
Lip, oral cavity, pharynx	605,098	6.2%
Esophagus	247,567	2.5%
Stomach	344,207	3.5%
Colon and rectum	2,386,178	24.5%
Liver	213,802	2.2%
Pancreas	516,375	5.3%
Larynx	209,337	2.1%
Trachea, lung, bronchus	3,002,109	30.8%
Cervix uteri	197,977	2.0%
Kidney and renal pelvis	759,615	7.8%
Urinary bladder	1,076,897	11.0%
Acute myeloid leukemia	188,936	1.9%

[31] 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. Available at www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/.
[32] Henley S.J., Singh S.D., King J, et al. Invasive cancer incidence and survival—United States, 2013. *MMWR* 2017.
[33] Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *Journal of the National Cancer Institute* 2011;103(24):1827–1839.
[34] 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.
[35] Levine B. What does the population attributable fraction mean? *Preventing Chronic Disease* 2007;4(1):A14.

SEER*Stat Item Name: **State race eth suppress**

Source of Standard: NPCR
Source Item Name: Derived from *Addr at DX - state* and state-level race or ethnicity reporting restrictions
Source Item Number: Derived from NAACCR 80 (Addr at Dx - state)

Description

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 combination.

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 race and ethnicity information cannot be reported at the state level.
- This variable should be used when conducting state-level analyses of race and ethnicity combinations. If you are conducting a state-level analysis of race- or ethnicity-only you should manually make restrictions in the SEER*Stat Selection tab.
- The following states have state-level race or ethnicity data presentation restrictions:
  - Data for American Indians and Alaska Natives cannot be displayed for Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York.
  - Data for Asians and Pacific Islanders cannot be displayed for Delaware, Kansas and Kentucky.
  - Hispanic ethnicity data cannot be displayed for Delaware, Kentucky, and Massachusetts.
  - Race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
- For more information, please refer to the *Race Recode for USCS*, *Origin recode NHIA (Hispanic, Non-Hisp)*, and *NHIA derived Hisp origin* variable descriptions in this document.

Values	Frequency	Percentage
All races	26,658,070	26.1%
White	22,679,353	22.2%
White Non-Hispanic	18,594,782	18.2%
White Hispanic	1,626,760	1.6%
Black	2,745,850	2.7%
Black Non-Hispanic	2,506,607	2.5%
Black Hispanic	47,713	0.0%
American Indian/Alaska Native	127,970	0.1%
AIAN Non-Hispanic	125,133	0.1%
AIAN Hispanic	5,867	0.0%
Asian/Pacific Islander	725,408	0.7%
API Non-Hispanic	687,694	0.7%
API Hispanic	11,282	0.0%
Hispanic	1,777,144	1.7%
Non-Hispanic	23,726,184	23.3%

Additional Resources

- NPCR www.cdc.gov/cancer/npcr
- SEER https://seer.cancer.gov
- U.S. Cancer Statistics Publication Standard www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm
- NAACCR www.naaccr.org/
- NAACCR data dictionary https://www.naaccr.org/data-standards-data-dictionary
- American College of Surgeons Commission on Cancer (CoC) Registry Manuals: *Facility Oncology Registry Data Standards* (FORDS) or *Registry Operations and Data Standards* (ROADS) www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals
- SEER site recode ICD-O-3/WHO 2008 https://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html
- ICCC site recode ICD-O-3/WHO 2008 https://seer.cancer.gov/iccc/iccc-who2008.html
- AYA site recode ICD-O-3/WHO 2008 https://seer.cancer.gov/ayarecode/
- Lymphoma subtype recode ICD-O-3/WHO 2008 https://seer.cancer.gov/lymphomarecode/
- ICD-O-3 http://apps.who.int/iris/bitstream/10665/96612/1/9789241548496_eng.pdf
- Collaborative Staging Manual http://cancerstaging.org/cstage/manuals.html
- Census www.census.gov

Abbreviations

AI/AN	American Indian or Alaska Native
A/PI	Asian or Pacific Islander
AYA	Adolescent and young adult
CCR	Central cancer registry
CNS	Central nervous system
CoC	Commission on Cancer
CS	Collaborative Stage
Dx	Diagnosis
ICCC	International Classification of Childhood Cancer
ICD-O-3	*International Classification of Diseases for Oncology*, Third Edition
NAACCR	North American Association of Central Cancer Registries
NAPIIA NAACCR	Asian/Pacific Islander identification algorithm
NHIA NAACCR	Hispanic identification algorithm
NOS	Not otherwise specified
NPCR	National Program of Cancer Registries
PRCDA	Purchased/Referred Care Service Delivery Area
SEER	Surveillance, Epidemiology, and End Results
USCS	U.S. Cancer Statistics
WHO	World Health Organization

Appendix A. NPCR – Indian Health Services (IHS) Linkage Schedule

All NPCR-funded registries link with the Indian Health Service every five years. The most recent linkage year was 2016.

All state central cancer registries with Purchase/Referred Care Delivery Area (PRCDA)[1] counties link with the Indian Health Service 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
- Washington
- Wisconsin
- Wyoming

[1]Purchased/Referred Care Delivery Area (PRCDA) was previously referred to as Contract Health Service Delivery Area (CHSDA).