National Program of Cancer Registries (NPCR) Public Use Research Data

Data Standards and Data Dictionary

November, 2015 Submission

Diagnosis Years 2001–2013

National Center for Chronic Disease Prevention and Health Promotion

Division of Cancer Prevention and Control

CS273827-A

Table of Contents

Message from NPCR Leadership 2
Overview of National Program of Cancer Registries 3
Variable List 7
Data Citation 8
Cautionary Notes 8
Checklist for NPCR Public Use Dataset Analyses 11
Variable Definition and Frequency 12
Additional resources 52
Abbreviations 53
Appendix A. Indian Health Services Linkage Schedule 54

Message from NPCR Leadership

November 29, 2016

We are pleased to share data from the National Program of Cancer Registries (NPCR) in this new format with the public health research community. This will be the first time a cancer dataset this large and representative will be publicly and freely available to researchers.

Conducting public health surveillance is a core function of CDC and one the agency is proud to lead. We invest important resources into various disease surveillance systems to help protect our nation against dangerous health threats, including cancer. NPCR is one of the most comprehensive and complex of CDC's disease surveillance systems and involves compiling and disseminating information on over 1.6 million cancers annually, representing 96% of all invasive cancers diagnosed in the United States. When NPCR data are combined with the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) data, 100% US population coverage is achieved.

NPCR was first authorized by the US Congress in 1992. Though CDC cooperative agreements, the program currently supports 45 states, the District of Columbia, Puerto Rico, and the US Pacific Island Jurisdictions. Many of our supported partners have been submitting data to CDC since the mid-1990's.

We encourage researchers to use our data to help inform scientific inquiries, programs, and policies. This data source is a product of tremendous efforts by reporting facility staff, cancer registrars, central cancer registry staff, and CDC NPCR staff and contractors. Most importantly, however, these data represent individuals who have been affected by a cancer diagnosis. Through use of this dataset, researchers can continue to have a positive impact on comprehensive cancer prevention and control and the care and quality of lives for those diagnosed with cancer.

Sincerely,

A Blythe Ryerson, MPH, PhD
Lead Epidemiologist, Cancer Surveillance Branch
Division of Cancer Prevention and Control
National Center for Chronic Disease Prevention and Health Promotion

Overview of CDC's National Program of Cancer Registries

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 45 states, the District of Columbia, Puerto Rico, and the U.S. Pacific Island Jurisdictions (see map below). These central cancer registries represent 96% of the United States population (2013), including 96% of whites, 98% of African Americans, 96% of Hispanics, 93% of American Indians and Alaska Natives (AI/AN), and 93% of Asians and Pacific Islanders (A/PI).

The cancer registries routinely collect data on patient demographics, primary tumor site, tumor morphology, stage at diagnosis, first course of treatment, and outcomes. Multiple medical facilities, such as hospitals, doctor's offices, pathology laboratories, and other treatment centers send demographic and clinical information related to persons with incident cancer to the central cancer registry, where the information is consolidated. Every year since 2000, the NPCR central cancer registries have been submitting relevant demographic and clinical information about each diagnosed cancer case to CDC. **None of the information submitted to CDC contains personally identifiable information about individual patients.**

CDC works closely with a variety of partners to deliver and manage this cancer surveillance system. One of CDC's most critical partners is the National Cancer Institute (NCI), which funds the Surveillance, Epidemiology, and End Results (SEER) Program. Together, CDC's NPCR and NCI's SEER programs cover the entire United States population. These combined data are the official source of federal statistics on cancer incidence and are presented in the United States Cancer Statistics (USCS) Incidence and Mortality Web-based Report.

This national coverage enables researchers, clinicians, policy makers, public health professionals, and members of the public to monitor the burden of cancer, evaluate the successes of programs, and identify additional needs for cancer prevention and control efforts at national, state, and local levels. At NCI's request, data from the five states supported solely by the SEER program are not included in this database.

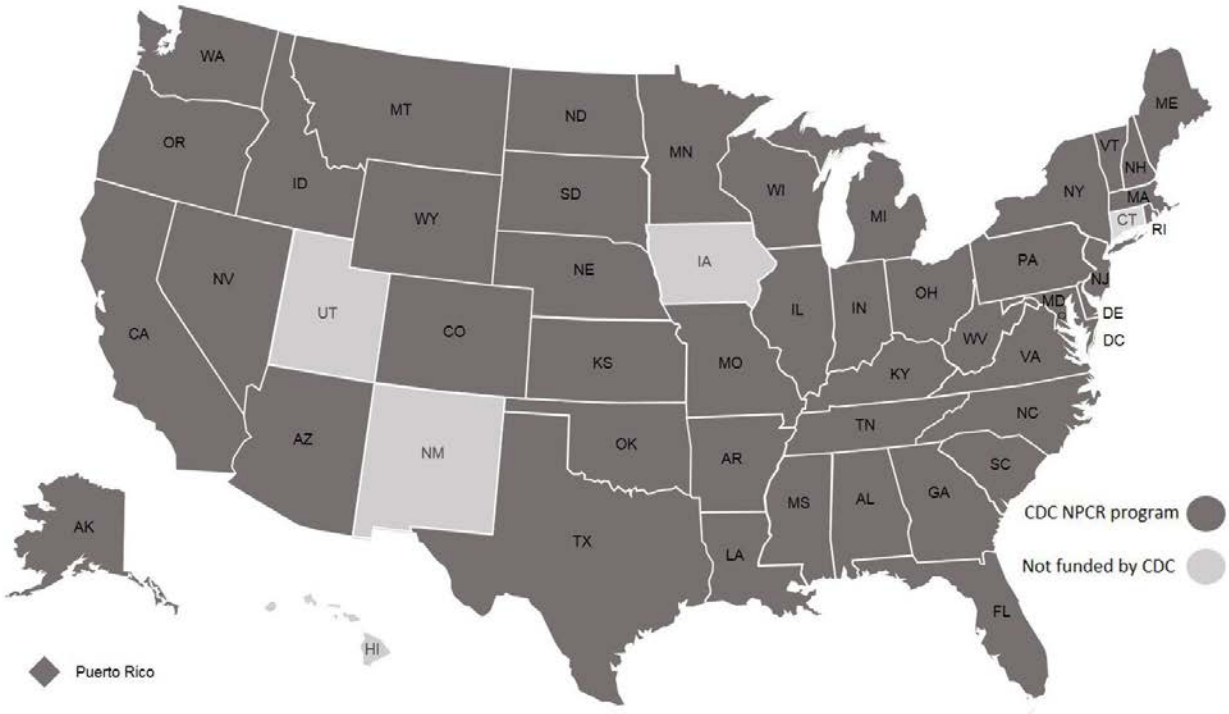

Figure 1. Central cancer registry programs funded by NPCR

NPCR Data Collection, Submission, Quality, and Coverage

The most current data come from the 2015 November NPCR submission, which covers incident cancer cases diagnosed in 2001 through 2013. Each year, NPCR-funded central cancer registries submit to CDC data on cancer diagnosed during the most recent year. In addition, data from previous years are updated with information from the newly submitted records. NPCR allows an interval of 23 months after the close of the diagnosis year for submission (by November 30, 2015 for 2013 data) to ensure case completeness and high quality. CDC supports the data collection and quality standards from the North American Association of Central Cancer Registries (NAACCR) consensus documents. During data collection, CDC also applies additional rigorous quality control edits, data completeness evaluations, and data quality assessments on all NPCR data. For a registry's data to be included in the NPCR public research data file, they must have met the quality and completeness criteria for publication in *United States Cancer Statistics (USCS)*:

- Case ascertainment greater than or equal to 90% (with a margin of error of plus or minus 5%).
- 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 2001–2013 Public Use Research Database

Two NPCR public use databases are available for researchers: the 2001–2013 database and the 2006–2013 database. This data standards document is specific to the 2001–2013 database.

The 2001–2013 database includes race and ethnicity variables, while the 2006–2013 database does not. The 2006–2013 database includes Puerto Rico data, while the 2001–2013 database does not.

- The 2001–2013 database's population denominators are race-specific, ethnicity-specific, and sex-specific county population estimates from the U.S. Census, modified by SEER and aggregated to the state and national levels.
- The 2006–2013 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 public research data.

For more detail on data availability by central cancer registry from 2001–2013, see Table 2 below.

Table 1. U.S. population coverage, NPCR 2015 submission 2001–2013 public use research database

Diagnosis year(s) [a]	% U.S. population covered in database
2001	90.3%
2002	90.2%
2003	95.9%
2004	95.9%
2005	95.9%
2006	95.9%
2007	95.9%
2008	95.9%
2009	95.9%
2010	95.9%
2011	95.0%
2012	95.0%
2013	95.0%
2001–2013	89.3%
2004–2013[b]	95.0%
2009–2013[c]	95.0%

[a] For the calculated percent population coverage for a range of years not shown in Table 1 (e.g., 2007–2012), please send a request to: npcrdata@cdc.gov.

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

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

Table 2. Central cancer registry data included in the 2001–2013 public use research database[a]

Registry	Year of Diagnosis												
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Alabama													
Alaska													
Arizona													
Arkansas													
California													
Colorado													
Delaware													
District of Columbia		NA											
Florida													
Georgia													
Idaho													
Illinois													
Indiana													
Kansas													
Kentucky													
Louisiana													
Maine													
Maryland													
Massachusetts													
Michigan													
Minnesota													
Mississippi	NS	NA											
Missouri													
Montana													
Nebraska													
Nevada											NA	NA	NA
New Hampshire													
New Jersey													
New York													
North Carolina													
North Dakota													
Ohio													
Oklahoma													
Oregon													
Pennsylvania													
Rhode Island													
South Carolina													
South Dakota													
Tennessee	NA	NA											
Texas													
Vermont													
Virginia	NA	NA											
Washington													
West Virginia													
Wisconsin													
Wyoming													

Shaded box: Data meet the *United States Cancer Statistics (USCS)* publication criteria and are available in the 2001-2013 public use research database.

NS: Data were not submitted and are not available in the 2001-2013 public use research database.

NA: Data did not meet USCS quality and completeness criteria for publication and are not available in the 2001-2013 public use research database.

[a] Puerto Rico and U.S. Pacific Island Jurisdiction data are not included in the 2001–2013 public use research database. Regardless of data completeness or quality, at NCI's request, data from states funded solely through the SEER program (Connecticut, Hawaii, Iowa, New Mexico, and Utah) are not included in this database.

Variable List

Table 3 shows all the variables available in the 2001-2013 public use research database.

Table 3. Variables in the 2001–2013 NPCR Public Use Research Database

SEER*Stat Category	SEER*Stat Variable Name
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
	USCS0113
	USCS0413
	Race recode for USCS
	USCS0913
	Region
	Stateraceethincl
	Origin recode NHIA (Hispanic, Non-Hisp)
Site and Morphology	Primary Site – labeled
	Histologic Type ICD-O-3
	Grade
	Diagnostic confirmation
	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
Extent of Disease – CS	Laterality
Multiple Primary Fields	Sequence number - central
Dates	Month of diagnosis
Other	Type of Reporting Source

Abbreviations used in variable names

Addr	Address
AYA	Adolescent and young adult
CS	Collaborative stage
Dx	Diagnosis
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
SS	Summary stage
USCS	United States Cancer Statistics
WHO	World Health Organization

Data Citation

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

- **For population coverage:** Data are from population-based registries that participate in the National Program of Cancer Registries 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 2001–2013 database:** National Program of Cancer Registries SEER*Stat Database: NPCR Incidence - Public Use Data, 2001–2013 - jbk 120616, United States Department of Health and Human Services, Centers for Disease Control and Prevention. Released January 2017, based on the November 2015 submission. Available at www.cdc.gov/cancer/npcr/public-use.

Cautionary Notes

Before using the database, analysts should read and understand the following nuances of the NPCR public use data. If you have questions regarding these notes, please contact CDC at npcrdata@cdc.gov.

Suppression Rules[1-2]

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.

The suppression rule is fewer than16 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 less than 16 for the time period. A count of less 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 less 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.

Note: As a further mechanism to protect data confidentiality, 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 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 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[3]

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 use 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 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[4-7]

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

http://www.naaccr.org/LinkClick.aspx?fileticket=U-3o31G2Lik%3d&tabid=126&mid=466.

Stage[8]

A merged variable, "Merged Summary Stage 2000," has been created to span two time periods when two different staging schemes were used. "SEER Summary Stage 2000" for diagnosis years 2001–2003 and the (or "Derived SEER Summary Stage 2000") for diagnosis years 2004–2013.

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 2013, then the stage at diagnosis is recorded using the "Derived SEER Summary Stage 2000" variable value.
- 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 2013 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

NPCR registries annually submit all eligible years of data to CDC. 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. 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.
2. Doyle P, Lane JI, Theeuwes JM, Zayatz LM. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies.* Amsterdam: Elsevier Science; 2001.
3. 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.
4. *International Classification of Diseases for Oncology.* Third Edition, First Revision. Geneva: World Health Organization, 2013.
5. Ruhl J, Adamo M, Dickie L. (January 2015). *Hematopoietic and Lymphoid Neoplasm Coding Manual*. National Cancer Institute, Bethesda, MD 20850-9765.
6. 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/seertools/hemelymph.
7. 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.
8. 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 NPCR Public Use Dataset Analyses

- ☐ If you are reporting state-specific ethnicity or race/ethnicity combinations, have you suppressed data from the registries that opted out of reporting these data items using the "Stateraceethincl" variable or manually in the SEER*Stat Selection tab?[1]
- ☐ If you are conducting a multiyear analysis and want to restrict it to the states that met reporting standards during each of the years, did you use variable USCS0913, USCS0413, or USCS1113 and also use the "Year of Diagnosis" variable to restrict to the corresponding year range on the SEER*Stat Selection tab?
  - This is important to do during a trend analysis, as the same states need to be included for each year being analyzed.
  - The "Year of Diagnosis" variable is used in combination with the predefined USCS variable to exclude the nonrelevant years. For example, if USCS0913 is used, then "Year of Diagnosis" should also be restricted to diagnosis years 2009–2013 in the SEER*Stat Selection tab.
  - The variable USCS0913 includes states meeting USCS publication criteria for diagnosis years 2009–2013, USCS0413 for diagnosis years 2004–2013, and USCS0113 for diagnosis years 2001–2013. If you would like to analyze a range of years other than those predefined variables, please contact CDC at npcrdata@cdc.gov and we will create a new variable for you.[2]
- ☐ 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)?[3]
- ☐ 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?[4]
- ☐ 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?[5]
- ☐ 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 Public Use Research Database?[6]

---

[1] See: "Stateraceethincl," "Race Recode," and "Origin recode NHIA" variable descriptions.

[2] See "USCS standard," "USCS0911," "USCS0411," and "USCS0111" variable descriptions.

[3] See Cautionary Notes/Primary Site Variables section.

[4] See "Diagnostic Confirmation" variable descriptions.

[5] See "Sex" variable description.

[6] See Data Citation section.

NPCR 2015 Submission 2001–2013 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.

- “Malignant Behavior” is a default selection for this database, as this restriction is used by CDC’s NPCR and NCI’s SEER programs for generating 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.

- All cases with an unknown age or with sex other than male or female have been excluded from this database and are unavailable. 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.

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	13,734	0.1%
01–04 years	45,353	0.2%
05–09 years	33,123	0.2%
10–14 years	39,033	0.2%
15–19 years	66,224	0.3%
20–24 years	105,262	0.5%
25–29 years	165,123	0.8%
30–34 years	255,934	1.3%
35–39 years	401,632	2.0%
40–44 years	709,242	3.6%
45–49 years	1,144,725	5.7%
50–54 years	1,682,564	8.4%
55–59 years	2,143,127	10.7%
60–64 years	2,471,065	12.4%
65–69 years	2,672,918	13.4%
70–74 years	2,531,680	12.7%
75–79 years	2,301,375	11.5%
80–84 years	1,751,516	8.8%
85+ years	1,406,668	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	9,919,882	49.7%
Female	10,020,416	50.3%

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

Values	Frequency	Percentage
2001	1,321,844	6.6%
2002	1,335,619	6.7%
2003	1,409,262	7.1%
2004	1,467,394	7.4%
2005	1,500,437	7.5%
2006	1,537,690	7.7%
2007	1,589,204	8.0%
2008	1,610,160	8.1%
2009	1,631,248	8.2%
2010	1,619,597	8.1%
2011	1,642,493	8.2%
2012	1,633,442	8.2%
2013	1,641,908	8.2%

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, if funded by NPCR, 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 http://www.naaccr.org/StandardsandRegistryOperations/VolumeII.aspx
  - FORDS variable "state at diagnosis" at https://www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals

Values	Frequency	Percentage
Alaska	35,433	0.2%
Alabama	331,541	1.7%
Arkansas	199,787	1.0%
Arizona	363,444	1.8%
California	2,180,601	10.9%
Colorado	284,277	1.4%
District of Columbia	37,070	0.2%
Delaware	70,813	0.4%
Florida	1,482,400	7.4%
Georgia	574,564	2.9%
Idaho	95,216	0.5%
Illinois	890,901	4.5%
Indiana	439,908	2.2%
Kansas	195,746	1.0%
Kentucky	338,471	1.7%
Louisiana	314,174	1.6%
Massachusetts	513,492	2.6%
Maryland	385,102	1.9%
Maine	116,107	0.6%
Michigan	752,394	3.8%
Minnesota	362,349	1.8%
Missouri	418,794	2.1%
Mississippi	170,934	0.9%
Montana	73,679	0.4%
North Carolina	636,150	3.2%
North Dakota	47,758	0.2%
Nebraska	124,535	0.6%
New Hampshire	105,386	0.5%

New Jersey	691,967	3.5%
Nevada	114,856	0.6%
New York	1,485,821	7.5%
Ohio	831,428	4.2%
Oklahoma	249,863	1.3%
Oregon	273,249	1.4%
Pennsylvania	1,062,595	5.3%
Rhode Island	86,375	0.4%
South Carolina	321,724	1.6%
South Dakota	56,963	0.3%
Tennessee	373,135	1.9%
Texas	1,316,748	6.6%
Virginia	425,554	2.1%
Vermont	50,228	0.3%
Washington	466,404	2.3%
Wisconsin	403,943	2.0%
West Virginia	153,709	0.8%
Wyoming	34,710	0.2%

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

Source of Standard: NPCR
Source Item Name: Derived from NAACCR's "State at diagnosis (CoC)" and USCS publication standard
Source Item Number: Derived from NAACCR's 80

Description

This variable indicates the NPCR-funded central cancer registries with cancer incidence data that are of high quality and meet the USCS 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 USCS standard for an individual diagnosis 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, USCS0913 (includes diagnosis years 2009–2013), USCS0413 (includes diagnosis years 2004–2013), or USCS0113 (includes diagnosis years 2001–2013).
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at npcrdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.
- The USCS publication standard is available at https://www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

No. of CCRs [a]	Year of diagnosis	Frequency	Percentage
43	2001	1,321,844	6.6%
42	2002	1,335,619	6.7%
46	2003	1,409,262	7.1%
46	2004	1,467,394	7.4%
46	2005	1,500,437	7.5%
46	2006	1,537,690	7.7%
46	2007	1,589,204	8.0%
46	2008	1,610,160	8.1%
46	2009	1,631,248	8.2%
46	2010	1,619,597	8.1%
45	2011	1,642,493	8.2%
45	2012	1,633,442	8.2%
45	2013	1,641,908	8.2%

[a] Refer to Table 2 for the list of central cancer registries included in each diagnosis year.

CCR central cancer registries

SEER*Stat Item Name: **USCS0113**

Source of Standard: NPCR
Source Item Name: Derived from NAACCR's "State at diagnosis (CoC)" and USCS publication standards
Source Item Number: Derived from NAACCR's 80

Description

This variable indicates whether NPCR-funded central cancer registries met the USCS publication standard for all cancer sites combined each year in 2001–2013. When using this variable, restrict the diagnosis years to 2001–2013. 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 2001–2013 data in the 2001–2013 database. Data from states that did not meet the USCS publication standard in any given year were excluded from the database for that year. Please refer to Table 2 for the list of states by year that were included and excluded by this variable.
- 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, USCS0913 (includes diagnosis years 2009–2013), USCS0413 (includes diagnosis years 2004–2013), or USCS0113 (includes diagnosis years 2001–2013).
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at npcrdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.
- The USCS publication standard is available at https://www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

Values	Frequency	Percentage
Does not meet USCS standard from 2001–2013	1,121,549	5.6%
Meets USCS standard from 2001–2013	18,818,749	94.4%

SEER*Stat Item Name: **USCS0413**

Source of Standard: NPCR
Source Item Name: Derived from NAACCR's "State at diagnosis (CoC)" and USCS publication standard
Source Item Number: Derived from NAACCR's 80

Description

This variable indicates whether NPCR-funded central cancer registries met the USCS publication standard for all cancer sites combined each year in 2004–2013. When using this variable, restrict the diagnosis years to 2004–2013. 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 2004–2013 data in the 2001–2013 database. Data from states that did not meet the USCS publication standard in any given year were excluded from the database for that year. Please refer to Table 2 for the list of states by year that were included and excluded by this variable.
- 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, USCS0913 (includes diagnosis years 2009–2013), USCS0413 (includes diagnosis years 2004–2013) or USCS0113 (includes diagnosis years 2001–2013).
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at npcrdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.
- The USCS publication standard is available at https://www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

Values	Frequency	Percentage
Does not meet USCS standard from 2004–2013	83,270	0.5%
Meets USCS standard from 2004–2013	15,790,303	99.5%

SEER*Stat Item Name: **USCS0913**

Source of Standard: NPCR
Source Item Name: Derived from NAACCR's "State at diagnosis (CoC)" and USCS publication standard
Source Item Number: Derived from NAACCR's 80

Description

This variable indicates whether NPCR-funded central cancer registries met the USCS publication standard for all cancer sites combined each year in 2009–2013. When using this variable, restrict the diagnosis years to 2009–2013. 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 2009–2013 data in the 2001–2013 database. Data from states that did not meet the USCS publication standard in any given year were excluded from the database for that year. Please refer to Table 2 for the list of states by year that were included and excluded by this variable.
- 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, USCS0913 (includes diagnosis years 2009–2013), USCS0413 (includes diagnosis years 2004–2013) or USCS0113 (includes diagnosis years 2001–2013).
- If you would like to analyze a range of years other than those predefined variables, please contact CDC at npcrdata@cdc.gov and we will create a new variable for you that can be imported into SEER*Stat.
- The USCS publication standard is available at https://www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm.

Values	Frequency	Percentage
Does not meet USCS standard from 2009–2013	25,217	0.3%
Meets USCS standard from 2009–2013	8,143,471	99.7%

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 USCS 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 USCS 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: http://www.cdc.gov/cancer/npcr/uscs/data/00_bias_correction.htm.
- If any state in a region has a case count of less than 16, then the case counts for U.S. Census regions cannot be presented.
- See https://www.census.gov/geo/reference/gtc/gtc_census_divreg.html for a list of states in each region.

Values	Frequency	Percentage
Northeast	4,111,971	20.6%
Midwest	4,524,719	22.7%
South	7,381,739	37.0%
West	3,921,869	19.7%

SEER*Stat Item Name: **Stateraceethincl**

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.

Considerations for use

- 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 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 and Kentucky.
  - Hispanic ethnicity data cannot be displayed for Delaware, Kentucky, Massachusetts, North Dakota, and Virginia.
  - Race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Pennsylvania.
- For more information, please refer to the “Race Recode for USCS” and “Origin recode NHIA (Hispanic, Non-Hisp)” variable descriptions in this document.

Values	Frequency	Percentage
Exclude state for race/ethnicity state-level analyses	5,723,118	28.7%
Include state for race/ethnicity state-level analyses	14,217,180	71.3%

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 2001–2013 public use database; it is not available in the 2006–2013 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 and Kentucky.
- 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 Contract Health Service Delivery Area (CHSDA) 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 2011 and 2016.
  - 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 analysis to 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 http://www.naaccr.org/StandardsandRegistryOperations/VolumeII.aspx
  - FORDS at https://www.facs.org/quality-programs/cancer/ncdb/registrymanuals/cocmanuals
  - SEER coding manuals http://seer.cancer.gov/tools/codingmanuals/historical.html

Values	Frequency	Percentage
White	16,999,888	85.3%
Black	2,080,423	10.4%
American Indian/Alaska Native	91,078	0.5%
Asian or Pacific Islander	473,435	2.4%
Other unspecified (1991+)	67,320	0.3%
Unknown	228,154	1.1%

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 2001–2013 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, Massachusetts, North Dakota, and Virginia.
  - Data for race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Pennsylvania.
- Blank values are allowed for states that chose not to include data for NHIA in this file.
- Additional information available at http://www.naaccr.org/LinkClick.aspx?fileticket=6E20OT41TcA%3d&tabid=118&mid=458.

Values	Frequency	Percentage
Non-Spanish-Hispanic-Latino	18,587,954	93.2%
Spanish-Hispanic-Latino	1,292,743	6.5%
Invalid Value(s)	59,601	0.3%

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)	19,598,723	98.3%
C80.9 (Unknown primary site)	341,575	1.7%

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 3–5 of the NAACCR 2010 Implementation Guidelines. http://www.naaccr.org/StandardsandRegistryOperations/ImplementationGuidelines.aspx.
- 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

Values	Frequency	Percentage
8000–9992	19,940,298	100%

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 (http://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."

Values	Frequency	Percentage
Well differentiated; Grade I	1,682,221	8.4%
Moderately differentiated; Grade II	5,000,044	25.1%
Poorly differentiated; Grade III	4,142,206	20.8%
Undifferentiated; anaplastic; Grade IV	629,794	3.2%
T-cell	65,704	0.3%
B-cell; pre-B; B-precursor	888,306	4.5%
Null cell; non T-non B	2,285	0.0%
NK cell; natural killer cell (1995+)	2,975	0.0%
Unknown	7,526,763	37.7%

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 that "diagnostic confirmation=microscopically confirmed" is selected.
- 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 https://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)	18,780,210	94.2%
Positive histology	18,050,557	90.5%
Positive exfoliative cytology, no positive histology	633,228	3.2%
Positive histology AND immunophenotyping AND/OR positive genetic studies	74,185	0.4%
Positive microscopic confirm, method not specified	22,240	0.1%
Positive laboratory test/marker study	85,596	0.4%
Direct visualization without microscopic confirmation	28,787	0.1%
Radiography without microscopic confirm	671,651	3.4%
Clinical diagnosis only	115,180	0.6%
Unknown	258,874	1.3%

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 is available at https://seer.cancer.gov/siterecode.

Values	Frequency	Percentage
All Sites (total)	**19,940,298**	
Oral Cavity and Pharynx	450,695	2.3%
Lip	27,881	0.1%
Tongue	129,152	0.6%
Salivary Gland	47,917	0.2%
Floor of Mouth	27,579	0.1%
Gum and Other Mouth	63,293	0.3%
Nasopharynx	21,455	0.1%
Tonsil	74,658	0.4%
Oropharynx	19,434	0.1%
Hypopharynx	28,754	0.1%
Other Oral Cavity and Pharynx	10,572	0.1%
Digestive System	3,377,316	16.9%
Esophagus	192,536	1.0%
Stomach	265,156	1.3%
Small Intestine	82,986	0.4%
Colon and Rectum	1,892,507	9.5%
Colon excluding Rectum	1,365,573	6.8%
Cecum	295,341	1.5%
Appendix	29,592	0.1%
Ascending Colon	261,739	1.3%
Hepatic Flexure	67,129	0.3%
Transverse Colon	123,003	0.6%
Splenic Flexure	42,592	0.2%
Descending Colon	80,807	0.4%
Sigmoid Colon	374,627	1.9%
Large Intestine, NOS	90,743	0.5%
Rectum and Rectosigmoid Junction	526,934	2.6%
Rectosigmoid Junction	140,086	0.7%
Rectum	386,848	1.9%
Anus, Anal Canal and Anorectum	76,433	0.4%

Values	Frequency	Percentage
Liver and Intrahepatic Bile Duct	253,957	1.3%
Liver	227,875	1.1%
Intrahepatic Bile Duct	26,082	0.1%
Gallbladder	45,136	0.2%
Other Biliary	66,157	0.3%
Pancreas	449,025	2.3%
Retroperitoneum	14,652	0.1%
Peritoneum, Omentum and Mesentery	23,152	0.1%
Other Digestive Organs	15,619	0.1%
Respiratory System	2,730,723	13.7%
Nose, Nasal Cavity and Middle Ear	27,236	0.1%
Larynx	164,524	0.8%
Lung and Bronchus	2,530,541	12.7%
Pleura	1,181	0.0%
Trachea, Mediastinum and Other Respiratory Organs	7,241	0.0%
Bones and Joints	35,136	0.2%
Soft Tissue including Heart	121,724	0.6%
Skin excluding Basal and Squamous	1,269,626	6.4%
Melanoma of the Skin	1,208,036	6.1%
Other Non-Epithelial Skin	61,590	0.3%
Breast (female and male combined)	3,236,959	16.2%
Female Genital System	1,062,781	5.3%
Cervix Uteri	155,617	0.8%
Corpus and Uterus, NOS	531,883	2.7%
Corpus Uteri	516,648	2.6%
Uterus, NOS	15,235	0.1%
Ovary	261,350	1.3%
Vagina	17,574	0.1%
Vulva	77,252	0.4%
Other Female Genital Organs	19,105	0.1%
Male Genital System	2,696,874	13.5%
Prostate	2,571,765	12.9%
Testis	98,516	0.5%
Penis	22,108	0.1%
Other Male Genital Organs	4,485	0.0%
Urinary System	1,476,190	7.4%
Urinary Bladder	821,670	4.1%
Kidney and Renal Pelvis	608,089	3.0%
Ureter	32,183	0.2%
Other Urinary Organs	14,248	0.1%
Eye and Orbit	36,556	0.2%
Brain and Other Nervous System	584,055	2.9%
Brain	270,350	1.4%
Cranial Nerves Other Nervous System	313,705	1.6%
Endocrine System	575,302	2.9%

Values	Frequency	Percentage
Thyroid	439,865	2.2%
Other Endocrine including Thymus	135,437	0.7%
Lymphoma	849,274	4.3%
Hodgkin Lymphoma	103,730	0.5%
Hodgkin – Nodal	100,539	0.5%
Hodgkin – Extranodal	3,191	0.0%
Non-Hodgkin Lymphoma	745,544	3.7%
NHL – Nodal	510,020	2.6%
NHL – Extranodal	235,524	1.2%
Myeloma	234,170	1.2%
Leukemia	497,533	2.5%
Lymphocytic Leukemia	247,594	1.2%
Acute Lymphocytic Leukemia	56,710	0.3%
Chronic Lymphocytic Leukemia	174,384	0.9%
Other Lymphocytic Leukemia	16,500	0.1%
Myeloid and Monocytic Leukemia	224,212	1.1%
Acute Myeloid Leukemia	145,332	0.7%
Acute Monocytic Leukemia	9,439	0.0%
Chronic Myeloid Leukemia	62,519	0.3%
Other Myeloid/Monocytic Leukemia	6,922	0.0%
Other Leukemia	25,727	0.1%
Other Acute Leukemia	9,960	0.0%
Aleukemic, Subleukemic and NOS	15,767	0.1%
Mesothelioma	38,890	0.2%
Kaposi Sarcoma	16,150	0.1%
Miscellaneous	650,344	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.
- Additional information is available at https://seer.cancer.gov/iccc/iccc3.html.

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

Values	Frequency	Percentage
I Leukemias, myeloproliferative & myelodysplastic diseases	46,936	23.8%
I(a) Lymphoid leukemias	33,560	17.0%
I(b) Acute myeloid leukemias	8,127	4.1%
I(c) Chronic myeloproliferative diseases	2,371	1.2%
I(d) Myelodysplastic syndrome and other myeloproliferative	1,556	0.8%
I(e) Unspecified and other specified leukemias	1,322	0.7%
II Lymphomas and reticuloendothelial neoplasms	26,434	13.4%
II(a) Hodgkin lymphomas	12,667	6.4%
II(b) Non-Hodgkin lymphomas (except Burkitt lymphoma)	8,859	4.5%
II(c) Burkitt lymphoma	2,445	1.2%
II(d) Miscellaneous lymphoreticular neoplasms	2,132	1.1%
II(e) Unspecified lymphomas	331	0.2%
III CNS and misc intracranial and intraspinal neoplasms	43,714	22.1%
III(a) Ependymomas and choroid plexus tumor	3,638	1.8%
III(b) Astrocytomas	16,616	8.4%
III(c) Intracranial and intraspinal embryonal tumors	6,255	3.2%
III(d) Other gliomas	5,531	2.8%
III(e) Other specified intracranial/intraspinal neoplasms	10,189	5.2%
III(f) Unspecified intracranial and intraspinal neoplasms	1,485	0.8%
IV Neuroblastoma and other peripheral nervous cell tumors	8,680	4.4%
IV(a) Neuroblastoma and ganglioneuroblastoma	8,448	4.3%
IV(b) Other peripheral nervous cell tumors	232	0.1%
V Retinoblastoma	3,210	1.6%
VI Renal tumors	7,009	3.5%
VI(a) Nephroblastoma and other nonepithelial renal tumors	6,322	3.2%
VI(b) Renal carcinomas	660	0.3%
VI(c) Unspecified malignant renal tumors	27	0.0%

Values	Frequency	Percentage
VII Hepatic tumors	2,263	1.1%
VII(a) Hepatoblastoma	1,648	0.8%
VII(b) Hepatic carcinomas	594	0.3%
VII(c) Unspecified malignant hepatic tumors	21	0.0%
VIII Malignant bone tumors	9,078	4.6%
VIII(a) Osteosarcomas	5,157	2.6%
VIII(b) Chondrosarcomas	342	0.2%
VII(c) Ewing tumor and related sarcomas of bone	2,976	1.5%
VIII(d) Other specified malignant bone tumors	421	0.2%
VIII(e) Unspecified malignant bone tumors	182	0.1%
IX Soft tissue and other extraosseous sarcomas	12,270	6.2%
IX(a) Rhabdomyosarcomas	4,850	2.5%
IX(b) Fibrosarcomas, peripheral nerve & other fibrous	1,321	0.7%
IX(c) Kaposi sarcoma	50	0.0%
IX(d) Other specified soft tissue sarcomas	4,746	2.4%
IX(e) Unspecified soft tissue sarcomas	1,303	0.7%
X Germ cell & trophoblastic tumors & neoplasms of gonads	11,690	5.9%
X(a) Intracranial & intraspinal germ cell tumors	2,043	1.0%
X(b) Extracranial & extragonadal germ cell tumors	1,457	0.7%
X(c) Malignant gonadal germ cell tumors	7,418	3.8%
X(d) Gonadal carcinomas	435	0.2%
X(e) Other and unspecified malignant gonadal tumors	337	0.2%
XI Other malignant epithelial neoplasms and melanomas	18,503	9.4%
XI(a) Adrenocortical carcinomas	222	0.1%
XI(b) Thyroid carcinomas	7,754	3.9%
XI(c) Nasopharyngeal carcinomas	612	0.3%
XI(d) Malignant melanomas	5,367	2.7%
XI(e) Skin carcinomas	92	0.0%
XI(f) Other and unspecified carcinomas	4,456	2.3%
XII Other and unspecified malignant neoplasms	798	0.4%
XII(a) Other specified malignant tumors	410	0.2%
XII(b) Other unspecified malignant tumors	388	0.2%
Not classified by ICCC or *in situ*	6,882	3.5%

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.

Values	Frequency	Percentage
I Leukemias, myeloproliferative & myelodysplastic diseases	46,936	23.8%
I(a) Lymphoid leukemias	33,560	17.0%
I(a.1) Precursor cell leukemias	32,567	16.5%
I(a.2) Mature B-cell leukemias	737	0.4%
I(a.3) Mature T-cell and NK cell leukemias	104	0.1%
I(a.4) Lymphoid leukemia, NOS	152	0.1%
I(b) Acute myeloid leukemias	8,127	4.1%
I(c) Chronic myeloproliferative diseases	2,371	1.2%
I(d) Myelodysplastic syndrome and other myeloproliferative	1,556	0.8%
I(e) Unspecified and other specified leukemias	1,322	0.7%
II Lymphomas and reticuloendothelial neoplasms	26,434	13.4%
II(a) Hodgkin lymphomas	12,667	6.4%
II(b) Non-Hodgkin lymphomas (except Burkitt lymphoma)	8,859	4.5%
II(b.1) Precursor cell lymphomas	2,519	1.3%
II(b.2) Mature B-cell lymphomas except Burkitt lymphoma	3,646	1.8%
II(b.3) Mature T-cell and NK-cell lymphomas	2,214	1.1%
II(b.4) Non-Hodgkin lymphomas, NOS	480	0.2%
II(c) Burkitt lymphoma	2,445	1.2%
II(d) Miscellaneous lymphoreticular neoplasms	2,132	1.1%
II(e) Unspecified lymphomas	331	0.2%
III CNS and misc intracranial and intraspinal neoplasms	43,714	22.1%
III(a) Ependymomas and choroid plexus tumor	3,638	1.8%
III(a.1) Ependymomas	2,821	1.4%
III(a.2) Choroid plexus tumor	817	0.4%
III(b) Astrocytomas	16,616	8.4%
III(c) Intracranial and intraspinal embryonal tumors	6,255	3.2%
III(c.1) Medulloblastomas	3,976	2.0%
III(c.2) PNET	1,439	0.7%
III(c.3) Medulloepithelioma	54	0.0%

Values	Frequency	Percentage
III(c.4) Atypical teratoid/rhabdoid tumor	786	0.4%
III(d) Other gliomas	5,531	2.8%
III(d.1) Oligodendrogliomas	713	0.4%
III(d.2) Mixed and unspecified gliomas	4,698	2.4%
III(d.3) Neuroepithelial glial tumors of uncertain orig	120	0.1%
III(e) Other specified intracranial/intraspinal neoplasms	10,189	5.2%
III(e.1) Pituitary adenomas and carcinomas	3,998	2.0%
III(e.2) Tumors of sellar region (craniopharyngiomas)	1,578	0.8%
III(e.3) Pineal parenchymal tumors	441	0.2%
III(e.4) Neuronal and mixed neuronal-glial tumors	3,035	1.5%
III(e.5) Meningiomas	1,137	0.6%
III(f) Unspecified intracranial and intraspinal neoplasms	1,485	0.8%
IV Neuroblastoma and other peripheral nervous cell tumors	8,680	4.4%
IV(a) Neuroblastoma and ganglioneuroblastoma	8,448	4.3%
IV(b) Other peripheral nervous cell tumors	232	0.1%
V Retinoblastoma	3,210	1.6%
VI Renal tumors	^1	^1
VI(a) Nephroblastoma and other nonepithelial renal tumors	^1	^1
VI(a.1) Nephroblastoma	5,945	3.0%
VI(a.2) Rhabdoid renal tumor	160	0.1%
VI(a.3) Kidney sarcomas	208	0.1%
VI(a.4) pPNET of kidney	^2	^2
VI(b) Renal carcinomas	660	0.3%
VI(c) Unspecified malignant renal tumors	27	0.0%
VII Hepatic tumors	2,263	1.1%
VII(a) Hepatoblastoma	1,648	0.8%
VII(b) Hepatic carcinomas	594	0.3%
VII(c) Unspecified malignant hepatic tumors	21	0.0%
VIII Malignant bone tumors	9,078	4.6%
VIII(a) Osteosarcomas	5,157	2.6%
VIII(b) Chondrosarcomas	342	0.2%
VIII(c) Ewing tumor and related sarcomas of bone	2,976	1.5%
VIII(c.1) Ewing tumor and Askin tumor of bone	2,863	1.4%
VIII(c.2) pPNET of bone	113	0.1%
VIII(d) Other specified malignant bone tumors	421	0.2%
VIII(d.1) Malignant fibrous neoplasms of bone	42	0.0%
VIII(d.2) Malignant chordomas	193	0.1%
VIII(d.3) Odontogenic malignant tumors	58	0.0%
VIII(d.4) Miscellaneous malignant bone tumors	128	0.1%
VIII(e) Unspecified malignant bone tumors	182	0.1%
IX Soft tissue and other extraosseous sarcomas	12,270	6.2%
IX(a) Rhabdomyosarcomas	4,850	2.5%
IX(b) Fibrosarcomas, peripheral nerve & other fibrous	1,321	0.7%
IX(b.1) Fibroblastic and myofibroblastic tumors	695	0.4%
IX(b.2) Nerve sheath tumors	608	0.3%
IX(b.3) Other fibromatous neoplasms	18	0.0%
IX(c) Kaposi sarcoma	50	0.0%
IX(d) Other specified soft tissue sarcomas	4,746	2.4%

Values	Frequency	Percentage
IX(d.1) Ewing tumor and Askin tumor of soft tissue	604	0.3%
IX(d.2) pPNET of soft tissue	294	0.1%
IX(d.3) Extrarenal rhabdoid tumor	239	0.1%
IX(d.4) Liposarcomas	259	0.1%
IX(d.5) Fibrohistiocytic tumors	1,195	0.6%
IX(d.6) Leiomyosarcomas	196	0.1%
IX(d.7) Synovial sarcomas	1,085	0.5%
IX(d.8) Blood vessel tumors	183	0.1%
IX(d.9) Osseous & chondromatous neoplasms of soft tissue	104	0.1%
IX(d.10) Alveolar soft parts sarcoma	154	0.1%
IX(d.11) Miscellaneous soft tissue sarcomas	433	0.2%
IX(e) Unspecified soft tissue sarcomas	1,303	0.7%
X Germ cell & trophoblastic tumors & neoplasms of gonads	^1	^1
X(a) Intracranial & intraspinal germ cell tumors	^1	^1
X(a.1) Intracranial & intraspinal germinomas	1,287	0.7%
X(a.2) Intracranial & intraspinal teratomas	506	0.3%
X(a.3) Intracranial & intraspinal embryonal carcinomas	22	0.0%
X(a.4) Intracranial & intraspinal yolk sac tumor	33	0.0%
X(a.5) Intracranial & intraspinal choriocarcinoma	^2	^2
X(a.6) Intracranial & intraspinal tumors of mixed forms	181	0.1%
X(b) Extracranial & extragonadal germ cell tumors	1,457	0.7%
X(b.1) Germinomas: extracranial/extragonadal	148	0.1%
X(b.2) Malignant teratomas: extracranial/extragonadal	565	0.3%
X(b.3) Embryonal carcinomas: extracranial/extragonadal	17	0.0%
X(b.4) Yolk sac tumor: extracranial/extragonadal	331	0.2%
X(b.5) Choriocarcinomas: extracranial/extragonadal	180	0.1%
X(b.6) Other mixed germ cell: extracranial/extragonadal	216	0.1%
X(c) Malignant gonadal germ cell tumors	7,418	3.8%
X(c.1) Malignant gonadal germinomas	1,546	0.8%
X(c.2) Malignant gonadal teratomas	1,306	0.7%
X(c.3) Gonadal embryonal carcinomas	763	0.4%
X(c.4) Gonadal yolk sac tumor	733	0.4%
X(c.5) Gonadal choriocarcinoma	73	0.0%
X(c.6) Malignant gonadal tumors of mixed forms	2,997	1.5%
X(d) Gonadal carcinomas	435	0.2%
X(e) Other and unspecified malignant gonadal tumors	337	0.2%
XI Other malignant epithelial neoplasms and melanomas	18,503	9.4%
XI(a) Adrenocortical carcinomas	222	0.1%
XI(b) Thyroid carcinomas	7,754	3.9%
XI(c) Nasopharyngeal carcinomas	612	0.3%
XI(d) Malignant melanomas	5,367	2.7%
XI(e) Skin carcinomas	92	0.0%
XI(f) Other and unspecified carcinomas	4,456	2.3%
XI(f.1) Carcinomas of salivary glands	905	0.5%
XI(f.2) Carcinomas of colon and rectum	599	0.3%
XI(f.3) Carcinomas of appendix	502	0.3%
XI(f.4) Carcinomas of lung	484	0.2%
XI(f.5) Carcinomas of thymus	67	0.0%
XI(f.6) Carcinomas of breast	213	0.1%

Values	Frequency	Percentage
XI(f.7) Carcinomas of cervix uteri	162	0.1%
XI(f.8) Carcinomas of bladder	306	0.2%
XI(f.9) Carcinomas of eye	31	0.0%
XI(f.10) Carcinomas of other specified sites	1,015	0.5%
XI(f.11) Carcinomas of unspecified site	172	0.1%
XII Other and unspecified malignant neoplasms	^[1]	^[1]
XII(a) Other specified malignant tumors	^[1]	^[1]
XII(a.1) Gastrointestinal stromal tumor	98	0.0%
XII(a.2) Pancreatoblastoma	34	0.0%
XII(a.3) Pulmonary blastoma and pleuropulmonary blastoma	187	0.1%
XII(a.4) Other complex mixed and stromal neoplasms	49	0.0%
XII(a.5) Mesothelioma	41	0.0%
XII(a.6) Other specified malignant tumors	^[2]	^[2]
XII(b) Other unspecified malignant tumors	388	0.2%
Not classified by ICCC or *in situ*	6,882	3.5%

[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, please visit https://seer.cancer.gov/ayarecode.

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

Values	Frequency	Percentage
1 Leukemias	20,980	6.2%
1.1 Acute lymphoid leukemia	8,208	2.4%
1.2 Acute myeloid leukemia	7,961	2.4%
1.3 Chronic myeloid leukemia	3,213	1.0%
1.4 Other and unspecified leukemia	1,598	0.5%
2 Lymphomas	49,958	14.8%
2.1 Non-Hodgkin lymphoma	19,251	5.7%
2.2 Hodgkin lymphoma	30,707	9.1%
3 CNS and Oth Intracranial and Intraspinal Neo (all behav)	29,956	8.9%
3.1. Astrocytoma	10,194	3.0%
3.1.1 Specified low-grade astrocytic tumors	4,092	1.2%
3.1.2 Glioblastoma and anaplastic astrocytoma	3,759	1.1%
3.1.3 Astrocytoma, NOS	2,343	0.7%
3.2 Other glioma	5,439	1.6%
3.3 Ependymoma	2,024	0.6%
3.4. Medulloblastoma and other PNET	1,811	0.5%
3.4.1 Medulloblastoma	999	0.3%
3.4.2 Supratentorial PNET	812	0.2%
3.5 Other specified intracranial and intraspinal neoplasms	8,825	2.6%
3.6 Unspecified intracranial and intraspinal neoplasms	1,663	0.5%
3.6.1 Unspec malignant intracranial and intraspinal neo	309	0.1%
3.6.2 Unspec ben/border intracran. and intraspin neo	1,354	0.4%
4 Osseous & Chondromatous Neoplasms	8,853	2.6%
4.1 Osteosarcoma	3,903	1.2%
4.2 Chondrosarcoma	1,172	0.3%
4.3 Ewing tumor	2,975	0.9%
4.4 Other specified and unspecified bone tumors	803	0.2%
5 Soft Tissue Sarcomas	14,569	4.3%

Values	Frequency	Percentage
5.1 Fibromatous neoplasms	3,666	1.1%
5.2 Rhabdomyosarcoma	1,675	0.5%
5.3 Other soft tissue sarcoma	9,228	2.7%
5.3.1 Specified soft tissue sarcoma	7,157	2.1%
5.3.1.1 Specified (excluding Kaposi sarcoma)	5,749	1.7%
5.3.1.2 Kaposi sarcoma	1,408	0.4%
5.3.2 Unspecified soft tissue sarcoma	2,071	0.6%
6 Germ Cell and Trophoblastic Neoplasms	40,007	11.9%
6.1 Germ cell and trophoblastic neoplasms of gonads	36,483	10.8%
6.2 Germ cell and trophoblastic neo of nongonadal sites	3,524	1.0%
6.2.1 Intracranial (all behaviors)	1,311	0.4%
6.2.2 Other nongonadal	2,213	0.7%
7 Melanoma and Skin Carcinomas	32,055	9.5%
7.1 Melanoma	31,789	9.4%
7.2 Skin carcinomas	266	0.1%
8 Carcinomas	97,098	28.8%
8.1 Thyroid carcinoma	41,367	12.3%
8.2 Other carcinoma of head and neck	5,663	1.7%
8.2.1 Nasopharyngeal carcinoma	1,057	0.3%
8.2.2 Other sites in lip, oral cavity and pharynx	4,060	1.2%
8.2.3 Nasal cav,mid ear,sinus,larynx,ill-def head/neck	546	0.2%
8.3 Carcinoma of trachea,bronchus, and lung	2,434	0.7%
8.4 Carcinoma of breast	12,313	3.7%
8.5 Carcinoma of genitourinary tract	19,933	5.9%
8.5.1 Carcinoma of kidney	4,000	1.2%
8.5.2 Carcinoma of bladder	1,749	0.5%
8.5.3 Carcinoma of gonads	3,068	0.9%
8.5.4 Carcinoma of cervix and uterus	10,570	3.1%
8.5.5 Carc of oth and ill-de41enitourinaryiourinary tract	546	0.2%
8.6 Carcinoma of gastrointestinal tract	13,529	4.0%
8.6.1 Carcinoma of colon and rectum	8,888	2.6%
8.6.2 Carcinoma of stomach	1,392	0.4%
8.6.3 Carcinoma of liver and intrahepatic bile ducts	1,367	0.4%
8.6.4 Carcinoma of pancreas	924	0.3%
8.6.5 Carc oth and ill-def sites, gastrointestinal tract	958	0.3%
8.7 Carcinoma of other and ill-def sites	1,859	0.6%
8.7.1 Adrenocortical carcinoma	287	0.1%
8.7.2 Carcinoma of other and ill-defined sites, NOS	1,572	0.5%
9 Miscellaneous specified neoplasms, NOS	7,545	2.2%
9.1 Other pediatric and embryonal tumors, NOS	787	0.2%
9.1.1 Wilms tumor	151	0.0%
9.1.2 Neuroblastoma	235	0.1%
9.1.3 Other pediatric and embryonal tumors, NOS	401	0.1%
9.2 Other specified and embryonal tumors, NOS	6,758	2.0%
9.2.1 Paraganglioma and glomus tumors	281	0.1%
9.2.2 Other specified gonadal tumors	571	0.2%
9.2.3 Myeloma, mast cell, misc lymphoreticular neo, NOS	1,140	0.3%
9.2.4 Other specified neoplasms, NOS	4,766	1.4%

Values	Frequency	Percentage
10 Unspecified Malignant Neoplasms	1,853	0.6%
Unclassified and Non-Malignant	33,735	10.0%

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.
- More information is available at https://seer.cancer.gov/lymphomarecode.

Values	Frequency	Percentage
Lymphoid Neoplasm	1,349,615	6.8%
1 Hodgkin Lymphoma	103,730	0.5%
1(a) Classical Hodgkin lymphoma	98,682	0.5%
1(a)1 Lymphocyte-rich/mixed cell/lymphocyte depleted	17,219	0.1%
1(a)1.1 Lymphocyte-rich	3,663	0.0%
1(a)1.2 Mixed cellularity	12,219	0.1%
1(a)1.3 Lymphocyte-depleted	1,337	0.0%
1(a)2 Nodular sclerosis	57,398	0.3%
1(a)3 Classical Hodgkin lymphoma, NOS	24,065	0.1%
1(b) Nodular lymphocyte predominant Hodgkin lymphoma	5,048	0.0%
2 Non-Hodgkin lymphoma	1,204,317	6.0%
2(a) Non-Hodgkin lymphoma, B-cell	1,106,902	5.6%
2(a)1 Precursor Non-Hodgkin lymphoma, B-cell	43,269	0.2%
2(a)2 Mature Non-Hodgkin lymphoma, B-cell	1,007,328	5.1%
2(a)2.1 Chronic/Sm/Prolymphocytic/Mantle B-cell NHL	249,869	1.3%
2(a)2.1.1 Chronic/Small lymphocytic leuk/lymph	218,555	1.1%
2(a)2.1.2 Prolymphocytic leukemia, B-cell	1,025	0.0%
2(a)2.1.3 Mantle-cell lymphoma	30,289	0.2%
2(a)2.2 Lymphoplasmacytic lymphoma/Waldenstrom	22,594	0.1%
2(a)2.2.1 Lymphoplasmacytic lymphoma	9,702	0.0%
2(a)2.2.2 Waldenstrom macroglubulinemia	12,892	0.1%
2(a)2.3 Diffuse large B-cell lymphoma (DLBCL)	260,353	1.3%
2(a)2.3.1 DLBCL, NOS	257,919	1.3%
2(a)2.3.2 Intravascular large B-cell lymphoma	450	0.0%
2(a)2.3.3 Primary effusion lymphoma	308	0.0%
2(a)2.3.4 Mediastinal large B-cell lymphoma	1,676	0.0%
2(a)2.4 Burkitt lymphoma/leukemia	15,090	0.1%

Values	Frequency	Percentage
2(a)2.5 Marginal-zone lymphoma (MZL)	68,500	0.3%
2(a)2.5.1 Splenic MZL	6,013	0.0%
2(a)2.5.2 Extranodal MZL, MALT type	41,334	0.2%
2(a)2.5.3 Nodal MZL	21,153	0.1%
2(a)2.6 Follicular lymphoma	144,318	0.7%
2(a)2.7 Hairy-cell leukemia	11,111	0.1%
2(a)2.8 Plasma cell neoplasms	235,298	1.2%
2(a)2.8.1 Plasmacytoma	15,931	0.1%
2(a)2.8.2 Multiple myeloma/plasma-cell leuk	219,367	1.1%
2(a)2.9 Heavy chain disease	195	0.0%
2(a)3 Non-Hodgkin lymphoma, B-cell, NOS	56,305	0.3%
2(b) Non-Hodgkin lymphoma, T-cell	77,218	0.4%
2(b)1 Precursor Non-Hodgkin lymphoma, T-cell	4,497	0.0%
2(b)2 Mature Non-Hodgkin lymphoma, T-cell	72,217	0.4%
2(b)2.1 Mycosis fungoides/Sezary syndrome	15,386	0.1%
2(b)2.1.1 Mycosis fungoides	14,820	0.1%
2(b)2.1.2 Sezary syndrome	566	0.0%
2(b)2.2 Peripheral T-cell lymphoma	43,244	0.2%
2(b)2.2.1 Peripheral T-cell lymphoma, NOS	15,066	0.1%
2(b)2.2.2 Angioimmunoblastic T-cell lymphoma	4,585	0.0%
2(b)2.2.3 Subcutan panniculitis-like T-cell lymph	371	0.0%
2(b)2.2.4 Anaplastic lar cell lymph, T-/Null-cell	10,114	0.1%
2(b)2.2.5 Hepatosplenic T-cell lymphoma	325	0.0%
2(b)2.2.6 Enteropathy-type T-cell lymphoma	501	0.0%
2(b)2.2.7 Cutaneous T-cell lymphoma, NOS	8,847	0.0%
2(b)2.2.8 Prim cutaneous anaplastic lar cell lymph	3,435	0.0%
2(b)2.3 Adult T-cell leukemia/lymphoma	7,231	0.0%
2(b)2.4 NK/T-cell lymph, nasal-type/aggres NK leuk	2,302	0.0%
2(b)2.5 T-cell large granular lymphocytic leukemia	2,745	0.0%
2(b)2.6 Prolymphocytic leukemia, T-cell	1,309	0.0%
2(b)3 Non-Hodgkin lymphoma, NOS, T-cell	504	0.0%
2(c) Non-Hodgkin lymphoma, unknown lineage	20,197	0.1%
2(c)1 Precursor lymphoblastic leuk/lymph, unk lineage	7,720	0.0%
2(c)2 Prolymphocytic leukemia, unknown lineage	569	0.0%
2(c)3 Non-Hodgkin lymphoma, NOS, unknown lineage	11,908	0.1%
3 Composite Hodgkin lymphoma and NHL	1,339	0.0%
4 Lymphoid neoplasm, NOS	40,229	0.2%
Unclassified	18,590,683	93.2%

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 (including “Malignant”, “Only malignant in ICD-O-3”, and “Only malignant 2010+”categories) 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 USCS official federal cancer statistics.
- For more information, please see SEER coding manual at http://seer.cancer.gov/icd-o-3.

Values	Frequency	Percentage
Benign	399,639	2.0%
Borderline malignancy	40,816	0.2%
In situ	1,332,496	6.7%
Malignant	17,884,739	89.7%
Only malignant in ICD-O-3	271,179	1.4%
Only malignant 2010+	11,429	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 and 759

Description

This is a merged stage variable created using two other variables: “SEER Summary Stage 2000,” which records stage from diagnosis years 2001–2003, and “Derived SS2000,” which records stage from diagnostic years 2004–2013. This stage variable can be used for diagnosis years 2001–2013.

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 2013, then the Derived Summary Stage 2000 (Derived SS2000) 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 was blank, but Summary Stage 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, please review https://cancerstaging.org/cstage/Pages/default.aspx.

Values	Frequency	Percentage
In situ	1,650,415	^[1]
Localized only	8,212,144	^[1]
Regional, direct extension only	1,326,534	^[1]
Regional, regional lymph nodes only	1,408,471	^[1]
Regional, direct extension and regional lymph nodes	826,850	^[1]
Regional, NOS	208,227	^[1]
Distant site(s)/node(s) involved	4,312,960	^[1]
Not applicable	440,170	^[1]
Unknown/unstaged/unspecified-	1,554,526	^[1]
Blank(s)	^[2]	^[2]

[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: **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 https://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	11,056,621	^[1]
Right - origin of primary	4,297,487	^[1]
Left - origin of primary	3,972,513	^[1]
Only one side - side unspecified	42,241	^[1]
Bilateral, single primary	157,689	^[1]
Paired site: midline tumor	29,787	^[1]
Paired site, but no information concerning laterality	383,959	^[1]
Invalid Value(s)	^[2]	^[2]

[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: **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	14,597,728	^[1]
1st of 2 or more primaries	1,575,507	^[1]
2nd of 2 or more primaries	2,777,093	^[1]
3rd of 3 or more primaries	451,694	^[1]
4th of 4 or more primaries	75,922	^[1]
5th of 5 or more primaries	14,789	^[1]
6th or more primaries[2]	6,695	^[1]
Only one state registry-defined neoplasm	422,638	^[1]
1st of 2 or more state registry-defined neoplasms	7,862	^[1]
2nd of 2 or more state registry-defined neoplasms	8,962	^[1]
3rd of 3 or more state registry-defined neoplasms	592	^[1]

Values	Frequency	Percentage
4th of 4 or more state registry-defined neoplasms	146	^[1]
5th of 5 or more state registry-defined neoplasms	74	^[1]
6th or more state registry-defined neoplasms[1]	30	^[1]
Carcinoma *in situ* of the Cervix diagnosed 1/1/1996 or later	^[3]	^[3]
Unknown sequence number - federally required *in situ* or malignant tumors	114	^[1]
Unknown sequence number - state registry-defined neoplasms	280	^[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: **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.

Values	Frequency	Percentage
January	1,762,231	8.8%
February	1,554,340	7.8%
March	1,702,487	8.5%
April	1,665,761	8.4%
May	1,684,518	8.4%
June	1,714,900	8.6%
July	1,629,758	8.2%
August	1,687,981	8.5%
September	1,595,031	8.0%
October	1,706,012	8.6%
November	1,568,364	7.9%
December	1,532,169	7.7%
Blank(s)	136,746	0.7%

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 case finding (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	17,500,897	87.8%
Radiation treatment or medical oncology center (2006+)	374,926	1.9%
Laboratory only (hospital or private)	494,329	2.5%
Physician's office/private medical practitioner (LMD)	864,068	4.3%
Nursing/convalescent home/hospice	25,335	0.1%
Other hospital outpatient unit or surgery center (2006+)	680,743	3.4%

Additional resources

- NPCR www.cdc.gov/cancer/npcr/
- USCS Publication Standard www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm
- NAACCR www.naaccr.org/
- NAACCR data dictionary www.naaccr.org/StandardsandRegistryOperations/VolumeII.aspx
- 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 https://seer.cancer.gov
- 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 / Alaska Native
A/PI	Asians and Pacific Islanders
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
SEER	Surveillance, Epidemiology, and End Results
SS	Summary Stage
USCS	United States Cancer Statistics
WHO	World Health Organization

Appendix A. 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 Contract Health Service Delivery Area (CHSDA) 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