U.S. Cancer Statistics
Restricted Access Data Set

Data Dictionary and Data Standards
2022 Data Submission
Released August 2023

Centers for Disease Control and Prevention
National Center for Chronic Disease Prevention & Health Promotion
Division of Cancer Prevention and Control
Cancer Surveillance Branch

National Cancer Institute
Division of Cancer Control and Population Sciences
Surveillance Research Program

Table of Contents

Overview 3

Available Data 4

Cautionary Notes 6

Population Denominator 12

Data Citation 13

References 14

Variable Descriptions 15

Appendix I 105

Appendix II 106

Appendix III 132

Overview

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, the U.S. Pacific Island Jurisdictions, and the U.S. Virgin Islands. Every year since 2000, the NPCR central cancer registries have been submitting relevant demographic and clinical information about each diagnosed cancer case to CDC.

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.

The programs' combined data are referred to as the U.S. Cancer Statistics (USCS) and they are the official source of federal statistics on cancer incidence. U.S. Cancer Statistics data are available to the public through various data products, including the U.S. Cancer Statistics Data Visualizations tool and public use database; external researchers have additional access to U.S. Cancer Statistics through the **U.S. Cancer Statistics Restricted Access Data Set**.

The U.S. Cancer Statistics Restricted Access Data Set is available to researchers in a SAS file at CDC's National Center for Health Statistics (NCHS) Research Data Centers (RDC) and Federal Statistical Researcher Data Centers. Researchers must complete a proposal and if the project is approved, adhere to RDC procedures for accessing the restricted data. Refer to the RDC website for details.

Available Data

This file documents the data items included in the **U.S. Cancer Statistics Restricted Access Data Set, 1998-2020.**

The purpose of this document is to define data standards for data items included in the U.S. Cancer Statistics Restricted Access Data Set (RADS) of the CDC's National Program of Cancer Registries (NPCR) Cancer Surveillance System (CSS) and NCI's Surveillance, Epidemiology, and End Results (SEER) Program. These variables are routinely collected through NPCR and SEER, and are defined by the North American Association of Central Cancer Registries (NAACCR). The following document describes the data items.

For all variables defined by NAACCR standards, abstractors use NAACCR's *Standards for Cancer Registries, Volume II: Data Standards and Data Dictionary,* in use for the given diagnosis year.

The data come from the 2022 NPCR-Cancer Surveillance System (NPCR-CSS) and SEER submissions.

- NPCR allowed an interval of 23 months after the close of the diagnosis year (data submission by November 30, 2022), and
- SEER allowed an interval of 22 months after the close of the diagnosis year (data submission by November 1, 2022).

For the list of central cancer registries available for analysis by year, please see Figure 1. The percent of cases covered by U.S. Cancer Statistics-eligible registries are listed below. If the year range you are analyzing is not listed, please e-mail CDC at uscsdata@cdc.gov and we will provide you the percentage.

- 1998-2020: 90.3%
- 1999-2020: 95.8%
- 2001-2020: 96.1%
- 2005-2020: 97.0%
- 2011-2020: 97.0%
- 2016-2020: 97.0%.

Figure 1. Central Cancer Registries Meeting U.S. Cancer Statistics Publication Criteria

2020: All registries met the publication criteria except Nevada and Indiana. Counts and rates cover approximately 97% of the U.S. population (48 U.S. States, D.C., and Puerto Rico).

2019: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2018: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2017: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2016: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2015: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2014: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2013: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2012: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2011: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2010: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2009: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2008: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2007: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2006: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2005: All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).

2004: All registries met the publication criteria (50 U.S. States and D.C.).

2003: All registries met the publication criteria (50 U.S. States and D.C.).

2002: All registries met the publication criteria except Mississippi. Counts and rates cover approximately 99% of the U.S. population (49 U.S. States and D.C.).

2001: All registries met the publication criteria except data are not available for Mississippi. Counts and rates cover approximately 99% of the U.S. population (49 U.S. States and D.C.).

2000: All registries met the publication criteria except data are not available for Mississippi and South Dakota. Counts and rates cover approximately 98.7% of the U.S. population (48 U.S. States and D.C.).

1999: All registries met the publication criteria except data are not available for Mississippi and South Dakota. Counts and rates cover approximately 98.7% of the U.S. population (48 U.S. States and D.C.).

1998: All registries met the publication criteria except Georgia and New Hampshire; data are not available for Mississippi, South Dakota, and Tennessee. Counts and rates cover approximately 93.4% of the U.S. population (45 U.S. States and D.C.).

In fall 2005, hurricanes Katrina and Rita hit the gulf coast and caused dramatic population shifts in the region. The US Census Bureau has provided estimates of the displaced populations within the four states of Alabama, Louisiana, Mississippi, and Texas. Use the adjusted US Census population estimates. Adjust county-level populations in the four hurricane-affected states to account for evacuations and that portion of the population be put into a “dummy” state (otherwise known as the KR area) for 2005.

Cautionary Notes

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

Central Cancer Registry Inclusion

Note that data from all registries are not represented each year. Data from each registry must meet eligibility criteria for inclusion in U.S. States Cancer Statistics to be included in this dataset and a state may be included for some years but not for all. See the U.S. Cancer Statistics publication criteria website for more information on the criteria.

This dataset includes Puerto Rico's data from diagnosis years 2005 to 2020. Note that Puerto Rico's 2017 incidence counts in the U.S. Cancer Statistics Data Visualizations tool and public use database are restricted to the first six months of reported data (January to June 2017). Data from July to December 2017 are excluded to account for the population shift that occurred due to Hurricane Maria. The population denominators were adjusted by dividing the U.S. Bureau of the Census's July 1, 2017 (vintage 2021) Puerto Rico population estimate in half.

In this Restricted Access Data Set, Puerto Rico's 2017 data are not restricted. Data for cases diagnosed from January to December 2017 are included. To calculate similar statistics as are found in the Data Visualizations tool or public use database, restrict analyses that include Puerto Rico 2017 cases using the *Date of Diagnosis* (SAS variable name: *I390_DateDx*) variable.

Four user-specified variables are included in the database: *uscs9820*, *uscs9920*, *uscs1120*, and *uscs1620*, for analyses using grouped years of data. These are particularly important for trend analyses, where the same states need to be included for each year under investigation. These user-specified variables contain all registries meeting U.S. Cancer Statistics criteria for all years included in the name of each variable (for example, *uscs1620* includes states that have data available for all five years, 2016 through 2020). Additionally, the variable, *USCS Standard*, is to be used for single year analyses.

Case Inclusions

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 *in situ*) central nervous system tumors are reported.
- *In situ* cancers of the urinary bladder are re-coded as invasive behavior and SEER Summary Stage *in situ* because the information needed to distinguish between *in situ* and invasive bladder cancers is not always available or reliable.[1]

Suppression Rules[2-3]

Effect of COVID-19 on Cancer Incidence Data for diagnosis year 2020

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

The 2022 data submission includes new cancer cases diagnosed in 2020, the first year of the COVID-19 pandemic. The COVID-19 pandemic disrupted health services, leading to delays and reductions in cancer screening, diagnosis, and reporting of data to some central cancer registries, which may have contributed to an observed decline in incidence for most cancer sites in 2020.

Complementary Cell Suppression

When analyzing data at the state or regional levels, suppress counts for national and regional data 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 applies when a single year or multiple years of data are being presented.

Suppressing fewer than 16 cases

The suppression rule is <16 cases for the time period based on rate stability. 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 or deaths, these statistics are not shown in tables and figures if the counts are fewer than 16 for the time period. A count of less than approximately 16 in a numerator results in a standard error of the rate that is approximately 25% or more as large as the rate itself. Equivalently, a count of less than approximately 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.

Reporting Delay[4]

Note that data are resubmitted by each NPCR and SEER registry each year. New cases are added each year to previous years resulting in a reporting delay. Cases may also be deleted from older years. Cases for certain primary sites e.g., melanoma and prostate, that are diagnosed on an out-patient basis can appear to be dropping in the most recent year.

Race and Ethnicity

Data 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 variables, *race_supp, raceeth_supp, nhia_supp*, can be used to restrict your analysis to the states that are eligible to be included in a state- or county-level analysis of race and ethnicity.

- If the variables, *race_supp, raceeth_supp, nhia_supp*, are not used and state- or county-level data are being reported, the following are the suppression restrictions for the 2022 submission data:

  The following states have data presentation restrictions —

  - Illinois, New Jersey, and New York – data for Hispanic and Non-Hispanic American Indian and Alaska Native persons cannot be displayed

  - Kansas – data for Hispanic and Non-Hispanic American Indian and Alaska Native, Asian and Pacific Islander, and Black persons cannot be displayed

  - North Dakota and Wisconsin – Hispanic ethnicity data alone or in combination with any race category cannot be displayed

Race Recode variable

The *Race recode (W, B, AI, API)* variable is created from *Race1*, *Race2*, and the Indian Health Service Link variable (*IHS Link*). Race/ethnicity starts as *Race1*. If *Race1* is white and *Race2* is a specified non-white race, then the value from *Race2* is used. After this check, if Race/ethnicity is still white and there is a positive *IHS Link*, then Race/Ethnicity is set to American Indian/Alaskan Native.

The *Race recode (W, B, AI, API)* variable contains an "other unspecified category". This group is treated as unknown race for the purpose of analyses as per the SEER documentation. Population data are not available for the other and unknown race categories.

Indian Health Service-linked American Indian/Alaska Natives (AI/AN) data

IHS provides medical services to AI/AN persons who are members of federally recognized tribes, estimated to be about 54% of the AI/AN population. To improve identification of AI/ANs, 33 NPCR registries with Purchase/Referred Care Delivery Area (PRCDA) counties in their state annually link with the IHS patient registration database to identify AI/ANs who had not been coded as AI/AN (see *IHS Link* variable description). All NPCR registries link every five years, with the most recent linkage occurring in 2021. SEER registries link their data annually, with the most recent linkage occurring among cases diagnosed from 1998-2020.

- 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.

The *race recode* variable contains Indian Health Service (IHS)-linked American Indian data.

Sex

When analyzing sex-specific cancers (such as prostate cancer or female breast cancer), the analysis should be limited to the appropriate sex to get the correct population denominator (e.g., only women or only men).

County

County data may be used only in approved analyses and in the following ways: a) used as a linkage variable only by the NCHS RDC analyst; b) included as a confounder or other control variable, but no data are presented by county; c) used in geographically aggregated form such as large metropolitan statistical areas (e.g., those with a population of 1 million or later), multi-county regions, or geographical areas (e.g., Appalachia or IHS Purchased/Referred Care Delivery Areas (PRCDA) counties). States are given the right to suppress county-specific data every submission year.

Stage

The variable, *Merged Summary Stage*, has been created to span four time periods when three different staging schemes were used. The coding logic for this merged variable is:

- For NPCR-registries—
    - If a case was diagnosed in 1998, 1999, or 2000, stage at diagnosis is recorded using the *SEER Summary Stage 1977* variable value.
    - If a case was diagnosed in 2001, 2002, 2003, 2016 or 2017, stage at diagnosis is recorded using the *SEER Summary Stage 2000* variable value.
    - If a case was diagnosed in or between 2004 and 2015, stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value. If the *Derived SEER Summary Stage 2000* variable is blank or unstaged, and the *SEER Summary Stage 2000* variable has a valid value, that value is used to populate the merged variable.
    - If a case was diagnosed in 2018 and after, stage at diagnosis is recorded using the *Summary Stage 2018* variable value.
- For SEER-only registries (Connecticut, Hawaii, Iowa, and New Mexico)—
    - If a case was diagnosed in 1998, 1999, or 2000, stage at diagnosis is recorded using the *SEER Summary Stage 1977* variable value.
    - If a case was diagnosed in 2001, 2002, or 2003, stage at diagnosis is recorded using the *SEER Summary Stage 2000* variable value.
    - If a case was diagnosed in or between 2004 and 2017, stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value.
    - If a case was diagnosed in 2018 and after, stage at diagnosis is recorded using the *Derived Summary Stage 2018* variable value.

Merged Summary Stage variable is recommended for analyses that included cases diagnosed between 2001 and 2020. If your analysis includes cases diagnosed between 1998 and 2000, review the NAACCR documentation on changes from Summary Stage 1977 and Summary Stage 2000 before using the *Merged Summary Stage 2000* variable.

Notes for users of this variable include—

- Due to changes made in the *Summary Stage 2018* Coding Manual, for cases diagnosed in 2018 and after
    - the category Regional, NOS (code 5) is no longer used.

    - There is an artificial increase in the category Regional by Direct Extension Only (code 2) for brain, CNS Other, and lymphoma cases. This is because "Regional, NOS" for these cases changed from code 5 to code 2.

- *Merged Summary Stage* data are not available for testis, leukemia, or myeloma cases.

If the *Merged Summary Stage* variable is not used, please note the following variable history:

- The individual Collaborative Stage data elements began to be collected in diagnosis year 2004.
    - If a case was diagnosed in 2018 and after, stage at diagnosis is recorded using the *Summary Stage 2018* variable value.
    - *Derived Summary Stage* variable is used with 2004-2015 NPCR cases and 2004-2017 SEER-only cases.
    - *Summary Stage 2000* is used with 2001-2003 and 2016-2017 NPCR cases; it should be used with 2001-2003 SEER-only cases.
    - *Summary Stage 1977* is used with 1998-2000 cases.
- For primary sites where the coding instructions changed to redistribute the percentage of cases coded as localized, regional, and distant, limit analysis to 2001 cases and forward. See the NAACCR "Site-Specific Comparison of Summary Stage 1977" for specific information.

Primary Site Variables[5]

- 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 variables named *Primary Site or Primary Site – labeled*. For more information on the SEER primary site recodes, see http://seer.cancer.gov/siterecode.

- If a user-defined primary site variable is created rather than using *Site recode ICD-O-3/WHO 2008*, exclude leukemias and lymphomas (9590-9992). Users may also want to break out Kaposi sarcoma (9140) and mesothelioma (9050-9055). For more information on the SEER primary site recode, see http://seer.cancer.gov/siterecode.

If the analysis requires certain subsites, use *Primary Site*, with the exclusions described above. The *Site recode ICD-O-3/WHO 2008* variable collapses all subsites into the group and, in certain instances, uses the histology to include all cases, e.g., melanoma, lymphoma, Kaposi sarcoma.).

Benign Central Nervous System (CNS) Tumors

Cancer registries began collecting information on nonmalignant brain and other nervous system tumors beginning with 2004 diagnoses. 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. SEER registries voluntarily agreed to incorporate registration of these tumors in their standard practices. Data for nonmalignant brain and other nervous system tumors were available from all registries contributing to this report.

Histology

For analyses that include histology, restrict the analyses to cases where *Diagnostic Confirmation* is "microscopically confirmed." This restriction will additionally exclude the death certificate only (DCO) cases.

Behavior

- ICD-O-2 behavior coding was used for cases diagnosed between 1992 and 2000 (1998-2000 in this data set). Cases diagnosed January 1, 2001 and after use the ICD-O-3 classification system. ICD-O-2 cases were converted to ICD-O-3 before they were submitted for this data set.
- If the analysis only includes cases diagnosed in 2001 or later, use the variable *Behavior code ICD-O-3*.
- If the analysis includes cases diagnosed in 1998, 1999, and/or 2000, and also spans diagnosis years ≥2001, then use the variable, *Behavior recode for analysis derived/WHO2008*. This variable reconciles the differences between ICD-O-2 and ICD-O-3. The ICD-O-3 manual, Appendix 6, has a complete list of behavior code changes.

Histologic Type ICD-O-3[6-10]

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 https://www.naaccr.org/implementation-guidelines/#Previous_Guidelines.

Population Denominator

In U.S. Cancer Statistics data products, the population estimates for the denominators of incidence and death rates are race-specific, ethnicity-specific, and sex-specific county population estimates aggregated to the state or metropolitan-area level. The county population estimates are

a slight modification of the annual time series of July 1 county population estimates (by age, sex, race, and Hispanic origin) produced under a collaborative arrangement between the U.S. Bureau of the Census (Census Bureau) and CDC's National Center for Health Statistics with support from NCI through an interagency agreement. Single year of age population estimates by county can be downloaded from NCI SEER's website. The methods used to create these estimates are described on NCI's Single Year of Age County Population Estimates website.

Data Citation

The following standard citations are to be used for all 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 or Surveillance, Epidemiology, and End Results Program and meet high-quality data criteria. These registries cover approximately [XX][i] % of the U.S. population.
- **For age-adjusted rates:** Rates are per 100,000 persons (or per 100,000 men or per 100,000 women, if sex-specific cancer) and are age-adjusted to the 2000 U.S. standard population (19 age groups – Census P25–1130).
- **For the U.S. Cancer Statistics Restricted Access Data Set:** National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: U.S. Cancer Statistics Restricted Access Data Set, November 2022 submission (1998-2020), United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released August 2023, based on November 2022 submissions.

[i] If the year range you are analyzing is not listed on page 4 of this document, please e-mail CDC at uscsdata@cdc.gov and we will provide you the percentage.

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. 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.
5. 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.
6. *International Classification of Diseases for Oncology.* Third Edition, First Revision. Geneva: World Health Organization, 2013.
7. Ruhl J, Adamo M, Dickie L. (January 2015). *Hematopoietic and Lymphoid Neoplasm Coding Manual*. National Cancer Institute, Bethesda, MD 20850-9765.
8. 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; https://seer.cancer.gov/tools/mphrules/.
9. Dickie L., Johnson, CH., Adams, S., Negoita, S. (May 2023). Solid Tumor Rules. National Cancer Institute, Rockville, MD 20850; https://seer.cancer.gov/tools/solidtumor/.
10. Surveillance, Epidemiology, and End Results Program. *Hematopoietic and Lymphoid Neoplasm Database*. Bethesda, MD: US Department of Health and Human Services, National Cancer Institute; https://seer.cancer.gov/seertools/hemelymph.

Variable Descriptions

Section: Demographic Data Items

Alternate Patient ID Number 19

Address at Diagnosis – State.....20

Address at Diagnosis – County.....21

USCS Standard 22

USCS9920.....23

USCS1620.....24

USCS9820.....25

USCS1120.....26

Address at Diagnosis – Census Region 27

Race 1.....28

Race 2.....29

Race recode (W, B, AI, API) 30

Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic) 32

Origin Recode NHIA (Hispanic, Non-Hisp) 34

IHS Link.....35

State Race Ethnicity Suppress variables.....37

Sex.....38

Age at Diagnosis.....39

Age Recode.....40

Birth Date.....41

Economic Status.....42

Rural-urban Continuum 2013……..44

Section: Cancer Identification Data Items

Sequence Number – Central 46

Year of Diagnosis 49

Date of Diagnosis.....50

Primary Site 51

Laterality.....52

Grade.....53

Grade Clinical 54

Grade Pathological.....55

Grade Post Therapy Path 56

Diagnostic Confirmation.....58

Type of Reporting Source.....59

Histologic Type ICD-O-3 60

Behavior Code ICD-O-3 61

Behavior Recode for Analysis Derived/WHO2008.....62

Site Recode ICD-O-3/WHO 2008 63

AYA Site Recode 2020 Revision 64

Lymphoid Neoplasm Recode 2021 Revision 65

ICCC Site Recode Extended ICD-O-3/WHO 2008.....66

Section: Stage/Prognostic Factors Data Items

SEER Summary Stage 2000 67

SEER Summary Stage 1977 68

Derived SS2000 69

Summary Stage 2018 71

Merged Summary Stage 72

CS Site-Specific Factor 1 73

Merged Estrogen Receptor 74

Merged Progesterone Receptor 75

Merged HER2 Receptor 76

Section: Treatment—First Course

RX Summ—Surgery Primary Site 77

Merged Radiation 78

Section: Over-ride flags

Over-ride Age/Site/Morph 79

Over-ride SeqNo/DxConf 81

Over-ride Site/Lat/Sequence Number 83

Over-ride Site/Type 85

Over-ride Histology 87

Over-ride Report Source 90

Over-ride Ill-define Site 92

Over-ride Leuk, Lymphoma 94

Over-ride Site/Behavior 96

Over-ride Site/Lat/Morph 98

Section: Risk Factor-Associated Cancers

Alcohol-related Cancers 100

HPV-related Cancers……..….101

Obesity-related Cancers…….…..102

Physical Activity-related Cancers……103

Tobacco-related Cancers…….……..…104

Section: Demographic Data Items
Alternate Patient ID Number

Alternate Patient ID Number

SAS Alternate Name	Length	Source of Standard
Random_ID	8	NAACCR Item #20

Description

Unique number assigned to an individual patient by the registry. A new unique number is assigned to each Patient ID Number prior to data release for confidentiality reasons. In combination with state at diagnosis, this should uniquely identify a person.

Codes

8-character code

Considerations for Use

None noted

Section: Demographic Data Items
Address at Diagnosis – State

Address at Diagnosis – State

SAS Alternate Name	Length	Source of Standard
I80_StateDx	2	NAACCR Item #80

Description

USPS abbreviation for the state, territory, commonwealth, or U.S. possession for the state/territory in which the patient resides at the time the reportable tumor is diagnosed. If the patient has multiple primaries, the state of residence may be different for each tumor.

Codes

Two letter USPS abbreviations for the 50 states, District of Columbia, and Puerto Rico. See Appendix II for variable coding.

Considerations for Use

The following states/diagnosis years are excluded from this file:

Diagnostic Year(s)	Exclusion state(s)
1998	Georgia, New Hampshire, Mississippi, South Dakota, Tennessee, Puerto Rico
1999	Mississippi, South Dakota, Puerto Rico
2000	Mississippi, South Dakota, Puerto Rico
2001	Mississippi, Puerto Rico
2002	Mississippi, Puerto Rico
2003	Puerto Rico
2004	Puerto Rico
2020	Nevada, Indiana

Section: Demographic Data Items
Address at Diagnosis – County

Address at Diagnosis – County

SAS Alternate Name	Length	Source of Standard
I89_CountyDxAnalysis	3	NAACCR Item #89

Description

Code for the county of the patient's residence at the time the tumor was diagnosed. For U.S. residents, standard codes are those of the FIPS publication "Counties and Equivalent Entities of the United States, Its Possessions, and Associated Areas." If the patient has multiple tumors, the county codes may be different for each tumor.

Codes

3-character value ranging from 001 to 998.

Codes need to be used in combination with I80_StateDx.

In addition to FIPS and Geocodes:
999 County unknown

Considerations for use

County data will be used only in approved analyses and in the following ways: a) used as a linkage variable (linkage to census data, for example) only by the NCHS RDC analyst; b) included as a confounder or other control variable, but no data are presented by county; c) used in geographically aggregated form such as large metropolitan statistical areas (e.g., those with a population of 1 million or larger), multi-county regions, or geographical areas (e.g., Appalachia or IHS Contract Health Services Delivery Areas (CHSDA) counties).

Kansas and Minnesota did not allow permission for their county data to be used. The *County at Diagnosis* variable for these states has been recoded to 999 for all diagnosis years. Virginia's data are suppressed in 2017-2020 (have been recoded to 999).

See the NAACCR data dictionary, Appendix A for standard FIPS county codes.

Section: Demographic Data Items
USCS Standard

USCS Standard

SAS Alternate Name	Length	Source of Standard
USCSSTD	1	NPCR

Description

This variable indicates the NPCR-funded 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.

Codes

1-character value

Note: USCSSTD = 1 if the state met USCS standards for specific DxYear. If the state didn't meet USCS standards for specific DxYear, this state is excluded from the U.S. Cancer Statistics Restricted Access Data Set.

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.
- 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, *USCS1620* (includes diagnosis years 2016-2020), *USCS1120* (includes diagnosis years 2011-2020), *USCS9920* (includes diagnosis years 1999-2020) or *USCS9820* (includes diagnosis years 1998–2020).
- The U.S. Cancer Statistics publication standard is available at https://www.cdc.gov/cancer/uscs/technical_notes/criteria/index.htm.

Section: Demographic Data Items
USCS9920

USCS9920

SAS Alternate Name	Length	Source of Standard
USCS9920	1	NPCR

Description

This variable indicates whether NPCR-funded central cancer registries met the U.S. Cancer Statistics publication standard for all cancer sites combined each year in 1999–2020. When using this variable, restrict the diagnosis years to 1999–2020.

Codes

1-character value

1 state met USCS standards for each DxYear in 1999-2020
0 state didn't meet USCS standards for each DxYear in 1999-2020

Considerations for use

- This variable is used for analysis of combined 1999–2020 data in the 1998–2020 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 the population coverage Excel file 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, *USCS1620* (includes diagnosis years 2016–2020), *USCS1120* (includes diagnosis years 2011–2020), *USCS9920* (includes diagnosis years 1999-2020) or *USCS9820* (includes diagnosis years 1998–2020).
- The U.S. Cancer Statistics publication standard is available at https://www.cdc.gov/cancer/uscs/technical_notes/criteria/index.htm.

Section: Demographic Data Items
USCS 1620

USCS1620

SAS Alternate Name	Length	Source of Standard
USCS1620	1	NPCR

Description

This variable indicates whether NPCR-funded central cancer registries met the U.S. Cancer Statistics publication standard for all cancer sites combined each year in 2016-2020 (the most recently submitted 5 years of data). When using this variable, restrict the diagnosis years to 2016-2020.

Codes

1-character value

1 state met USCS standards for each DxYear in 2016-2020
0 state didn't meet USCS standards for each DxYear in 2016-2020

Considerations for use

- This variable is used for analysis of combined 2016-2020 data in the 1998–2020 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 the population coverage Excel file 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, *USCS1620* (includes diagnosis years 2016–2020), *USCS1120* (includes diagnosis years 2011–2020), *USCS9920* (includes diagnosis years 1999-2020) or *USCS9820* (includes diagnosis years 1998–2020).
- The U.S. Cancer Statistics publication standard is available at https://www.cdc.gov/cancer/uscs/technical_notes/criteria/index.htm.

Section: Demographic Data Items
USCS9820

USCS9820

SAS Alternate Name	Length	Source of Standard
USCS9820	1	NPCR

Description

This variable indicates whether NPCR-funded central cancer registries met the U.S. Cancer Statistics publication standard for all cancer sites combined each year in 1998–2020. When using this variable, restrict the diagnosis years to 1998–2020.

Codes

1-character value

1 state met USCS standards for each DxYear in 1998-2020
0 state didn't meet USCS standards for each DxYear in 1998-2020

Considerations for use

- This variable is used for analysis of combined 1998–2020 data in the 1998–2020 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 the population coverage Excel file 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, *USCS1620* (includes diagnosis years 2016–2020), *USCS1120* (includes diagnosis years 2011–2020), *USCS9920* (includes diagnosis years 1999-2020) or *USCS9820* (includes diagnosis years 1998–2020).
- The U.S. Cancer Statistics publication standard is available at https://www.cdc.gov/cancer/uscs/technical_notes/criteria/index.htm.

Section: Demographic Data Items
USCS1120

USCS1120

SAS Alternate Name	Length	Source of Standard
USCS1120	1	NPCR

Description

This variable indicates whether NPCR-funded central cancer registries met the U.S. Cancer Statistics publication standard for all cancer sites combined each year in 2011-2020 (the most recently submitted 10 years of data). When using this variable, restrict the diagnosis years to 2011-2020.

Codes

1-character value

1	state met USCS standards for each DxYear in 2011-2020
0	state didn't meet USCS standards for each DxYear in 2011-2020

Considerations for use

- This variable is used for analysis of combined 2011-2020 data in the 1998–2020 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 the population coverage Excel file 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, *USCS1620* (includes diagnosis years 2016–2020), *USCS1120* (includes diagnosis years 2011–2020), *USCS9920* (includes diagnosis years 1999-2020) or *USCS9820* (includes diagnosis years 1998–2020).
- The U.S. Cancer Statistics publication standard is available at https://www.cdc.gov/cancer/uscs/technical_notes/criteria/index.htm.

Section: Demographic Data Items
Address at Diagnosis – Census Region

Address at Diagnosis – Census Region

SAS Alternate Name	Length	Source of Standard
CENSUS_REGION	9	Derived based upon NAACCR Item #80

Description

The region where the patient lived at diagnosis.

The NAACCR data item Address at Diagnosis—State [80] is recoded into one of the four *Census regions*, the same definition used for region in United States Cancer Statistics. Reference us_regdiv.pdf (census.gov) for a list of states for each region.

Codes

Midwest
Northeast
South
West

Considerations for Use

This data item is not available for Puerto Rico.

Section: Demographic Data Items

Race 1

Race 1

SAS Alternate Name	Length	Source of Standard
I160_Race1	2	NAACCR Item #160

Description

Code for the patient's race. Race is coded separately from Spanish/Hispanic Origin [190]. All tumors for the same patient should have the same race codes. If the patient is multiracial, a second race is coded in the data item RACE 2 [161]. For coding instructions and race code history see the current *SEER Program Coding and Staging Manual*. Reference to Census 2000 definitions for ethnicity and race: http://www.census.gov/prod/cen2000/doc/sf2.pdf

Codes

01 White
02 Black
03 American Indian, Aleutian, or Alaska Native (includes all indigenous populations of the Western hemisphere)
04 Chinese
05 Japanese
06 Filipino
07 Hawaiian
08 Korean
10 Vietnamese
11 Laotian
12 Hmong
13 Kampuchean (Cambodian)
14 Thai
15 Asian Indian or Pakistani, NOS (code 09 prior to Version 12)
16 Asian Indian
17 Pakistani
20 Micronesian, NOS
21 Chamorro/Chamoru
22 Guamanian, NOS
25 Polynesian, NOS
26 Tahitian
27 Samoan
28 Tongan
30 Melanesian, NOS
31 Fiji Islander
32 New Guinean
96 Other Asian, including Asian, NOS and Oriental, NOS
97 Pacific Islander, NOS
98 Other
99 Unknown

Considerations for Use

Population data are not available for this variable. This data item is not available for Puerto Rico. For age-adjusted rates by race, use *Race_recode_W_B_AI_API*.

The following states have state-level race data presentation restrictions:

- Illinois, New Jersey, and New York – data for Hispanic and Non-Hispanic American Indian and Alaska Native persons cannot be displayed
- Kansas – data for Hispanic and Non-Hispanic American Indian and Alaska Native, Asian and Pacific Islander, and Black persons cannot be displayed
- North Dakota and Wisconsin – Hispanic ethnicity data alone or in combination with any race category cannot be displayed

Section: Demographic Data Items
Race 2

Race 2

SAS Alternate Name	Length	Source of Standard
I161_Race2	2	NAACCR Item #161

Description

Code for the patient's race. Race is coded separately from Spanish/Hispanic Origin [190]. All tumors for the same patient should have the same race codes. If the patient is multiracial, the second race is coded in this data item. For coding instructions and race code history see the current *SEER Program Coding and Staging Manual*. Reference to Census 2000 definitions for ethnicity and race: http://www.census.gov/prod/cen2000/doc/sf2.pdf

Codes

01 White
02 Black
03 American Indian, Aleutian, or Alaska Native (includes all indigenous populations of the Western hemisphere)
04 Chinese
05 Japanese
06 Filipino
07 Hawaiian
08 Korean
10 Vietnamese
11 Laotian
12 Hmong
13 Kampuchean (Cambodian)
14 Thai
15 Asian Indian or Pakistani, NOS (code 09 prior to Version 12)
16 Asian Indian
17 Pakistani
20 Micronesian, NOS
21 Chamorro/Chamoru
22 Guamanian, NOS
25 Polynesian, NOS
26 Tahitian
27 Samoan
28 Tongan
30 Melanesian, NOS
31 Fiji Islander
32 New Guinean
96 Other Asian, including Asian, NOS and Oriental, NOS
97 Pacific Islander, NOS
98 Other
99 Unknown

Considerations for Use

Population data are not available for this variable. This data item is not available for Puerto Rico. For age-adjusted rates by race, use *Race_recode_W_B_AI_API*.

The following states have state-level race data presentation restrictions:

- Illinois, New Jersey, and New York – data for Hispanic and Non-Hispanic American Indian and Alaska Native persons cannot be displayed
- Kansas – data for Hispanic and Non-Hispanic American Indian and Alaska Native, Asian and Pacific Islander, and Black persons cannot be displayed
- North Dakota and Wisconsin – Hispanic ethnicity data alone or in combination with any race category cannot be displayed

Section: Demographic Data Items
Race recode (W, B, AI, API)

Race Recode (W, B, AI, API)

SAS Alternate Name	Source of Standard
Race_recode_W_B_AI_API	Derived based upon NAACCR Items #160 and #161

Description

This variable indicates the derived code for the patient's race. Race is coded separately from Hispanic ethnicity.
Data quality checks code a non-White race before a White race. This variable is created using NAACCR variables Race 1 and the Indian Health Service (IHS) link. If Race 1 is White and there is a positive IHS Link, then Race/Ethnicity is set to American Indian/Alaskan Native (AI/AN).

Considerations for Use

- This data item is not available for Puerto Rico.
- 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—
    - Illinois, New Jersey, and New York – data for Hispanic and Non-Hispanic American Indian and Alaska Native persons cannot be displayed
    - Kansas – data for Hispanic and Non-Hispanic American Indian and Alaska Native, Asian and Pacific Islander, and Black persons cannot be displayed
    - North Dakota and Wisconsin – Hispanic ethnicity data alone or in combination with any race category cannot be displayed

  Race is defined by specific physical, hereditary, and cultural traditions or origins, not necessarily by birthplace, place of residence, or citizenship. "Origin" is defined by the U.S. Census Bureau as the heritage, nationality group, lineage, or in some cases, the country of birth of the person or the person's parents or ancestors before their arrival in the United States. As a standard practice, central cancer registries classify race as coded in the medical record. To address AI/AN misclassification in cancer registry data, registries supported by CDC's National Program of Cancer Registries Program (NPCR) and the National Cancer Institute's Surveillance Epidemiology End Results (SEER) Program link their central cancer registry data to the Indian Health Service (IHS) administrative records database.
    - SEER registries link their data annually, with the most recent linkage occurring among cases diagnosed from 1998 to 2020. Annually, 33 NPCR registries with Purchase/Referred Care Delivery Area (PRCDA) counties in their state link their data. All NPCR registries link every five years, with the most recent linkage occurring in 2021.

    - Although the linkage with IHS does not completely resolve the classification of race for AI/AN cases, it helps provide a more comprehensive and accurate picture of the cancer burden in this population.
    - When interpreting data results, be aware that AI/AN populations may be undercounted during years when linkages did not occur in all NPCR registries.
    - If a project is looking specifically at AI/AN populations, analysts may consider restricting the NPCR states included in the analysis to NPCR registries that conduct annual IHS linkages.
- In all separate records of tumors for the same patient, the patient has the same race code.
- This variable contains "other unspecified" and "unknown" categories. These groups are coded as "unknown race" for the purpose of analyses as specified in the SEER documentation. Population data are not available for the "other race" and "unknown race" categories.
- For further information on creating this variable, see the SAS statements in Appendix I.

Codes

See Appendix II for variable coding.

Considerations for Use

This data item is not available for Puerto Rico.

Section: Demographic Data Items
Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)

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

SAS Alternate Name	Source of Standard
Race_and_origin_recode	Derived based upon NAACCR Items #160, #161, #191, and #192

Description

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

Considerations for Use

- This data item is not available for Puerto Rico.
- 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—
    - Illinois, New Jersey, and New York – data for Hispanic and Non-Hispanic American Indian and Alaska Native persons cannot be displayed
    - Kansas – data for Hispanic and Non-Hispanic American Indian and Alaska Native, Asian and Pacific Islander, and Black persons cannot be displayed
    - North Dakota and Wisconsin – Hispanic ethnicity data alone or in combination with any race category cannot be displayed

  Race is defined by specific physical, hereditary, and cultural traditions or origins, not necessarily by birthplace, place of residence, or citizenship. "Origin" is defined by the U.S. Census Bureau as the heritage, nationality group, lineage, or in some cases, the country of birth of the person or the person's parents or ancestors before their arrival in the United States. As a standard practice, central cancer registries classify race as coded in the medical record. To address AI/AN misclassification in cancer registry data, registries supported by CDC's National Program of Cancer Registries Program (NPCR) and the National Cancer Institute's Surveillance Epidemiology End Results (SEER) Program link their central cancer registry data to the Indian Health Service (IHS) administrative records database.
    - SEER registries link their data annually, with the most recent linkage occurring among cases diagnosed from 1998 to 2020. Annually, 33 NPCR registries with Purchase/Referred Care Delivery Area (PRCDA) counties in their state link their data. All NPCR registries link every five years, with the most recent linkage occurring in 2021.
    - Although the linkage with IHS does not completely resolve the classification of race for AI/AN cases, it helps provide a more comprehensive and accurate picture of the cancer burden in this population.
    - When interpreting data results, be aware that AI/AN populations may be undercounted during years when linkages did not occur in all NPCR registries.

    - If a project is looking specifically at AI/AN populations, analysts may consider restricting the NPCR states included in the analysis to NPCR registries that conduct annual IHS linkages.
- In all separate records of tumors for the same patient, the patient has the same race code.
- This variable contains "other unspecified" and "unknown" categories. These groups are coded as "unknown race" for the purpose of analyses as specified in the SEER documentation. Population data are not available for the "other race" and "unknown race" categories.

Codes

See Appendix II for variable coding.

Considerations for Use

This data item is not available for Puerto Rico.

Section: Demographic Data Items
Origin Recode NHIA (Hispanic, Non-Hisp)

Origin Recode NHIA (Hispanic, Non-Hisp)

SAS Alternate Name	Length	Source of Standard
Origin Recode NHIA	1	NAACCR Item #191

Description

The NAACCR Hispanic Identification Algorithm (NHIA) uses a combination of standard variables to directly or indirectly classify cases as Hispanic for analytic purposes. It is possible to separate Hispanic ancestral subgroups (e.g., Mexican) when indirect assignment results from birthplace information but not from surname match. The algorithm uses the following standard variables: Spanish/Hispanic Origin [190], Name--Last [2230], Name--Maiden [2390], Birthplace [250], Race 1 [160], IHS Link [192], and Sex [220].

Codes

0=Non-Spanish-Hispanic-Latino
1=Spanish-Hispanic-Latino
3=Unknown / invalid
9=Suppressed, North Dakota and Wisconsin

Note: Code 3 for Puerto Rico only

Considerations for Use

This data item is not available for Puerto Rico.

The following states have state-level race or ethnicity data presentation restrictions:

- Illinois, New Jersey, and New York – data for Hispanic and Non-Hispanic American Indian and Alaska Native persons cannot be displayed
- Kansas – data for Hispanic and Non-Hispanic American Indian and Alaska Native, Asian and Pacific Islander, and Black persons cannot be displayed
- North Dakota and Wisconsin – Hispanic ethnicity data alone or in combination with any race category cannot be displayed

Section: Demographic Data Items
IHS Link

IHS Link

SAS Alternate Name	Length	Source of Standard
I192_IHS	1	NAACCR Item #192

Description

This variable captures the results of the linkage of the registry database with the Indian Health Service patient registration database. The IHS linkage identifies cancer cases among American Indians/Alaskan Natives who were misclassified as non-Indian in the registry database in order to improve the quality of cancer surveillance data on American Indians/Alaskan Natives in individual registries and in all registries as a whole. The goal is to improve cancer incidence data for American Indians/Alaskan Natives in the United States Cancer Statistics by use of this variable as well as the race variable.

Codes

0	Record sent for linkage, no IHS match
1	Record sent for linkage, IHS match
Blank	Record not sent for linkage or linkage result pending

Considerations for Use

- **This variable includes only count data**. Rates cannot be calculated using this variable as no population data are associated it.
- IHS provides medical services to American Indians and Alaska Natives (AI/ANs) who are members of federally recognized tribes, estimated to be about 54% of the AI/AN population. To improve identification of AI/ANs, 33 NPCR registries with Purchase/Referred Care Delivery Area (PRCDA) 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 table below). All NPCR registries link every five years, with the most recent linkage occurring in 2021. SEER registries link their data annually, with the most recent linkage occurring among cases diagnosed from 1998 to 2020.
    - 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 the table below for the list of these states.
- Those registries **not** included in the table below may elect to link with IHS annually, but are required to link every five years. Blank values are allowed for states without PRCDA counties that chose to not link with IHS annually or chose to not include data for American Indians/Alaskan Natives in this file.

- Population data are not available for this variable. For age-adjusted rates by race, use the variable *Race recode (W, B, AI, API*).

Table. NPCR Registries with one or more IHS PRCDA county.

Alabama	Massachusetts	Oregon
Alaska	Michigan	Pennsylvania
Arizona	Minnesota	Rhode Island
California	Mississippi	South Carolina
Colorado	Montana	South Dakota
Florida	Nebraska	Texas
Idaho	Nevada	Utah
Indiana	New York	Virginia
Kansas	North Carolina	Washington
Louisiana	North Dakota	Wisconsin
Maine	Oklahoma	Wyoming

Section: Demographic Data Items
State race ethnicity suppress

State race ethnicity suppress

SAS Alternate Name	Length	Source of Standard
race_supp	1	Derived based on NAACCR Items #80, #160, #161, #191 and #192
raceeth_supp	1	
nhia_supp	1	

Description

Those variables were created specifically for this dataset. It provides the selection of states that are eligible to be included in a state-level analysis of race and ethnicity.

Codes

See Appendix III for variables coding.

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.
- This variable is used when conducting state-level analyses of race and ethnicity combinations.
- The following states have state-level race or ethnicity data presentation restrictions:
  - Illinois, New Jersey, and New York – data for Hispanic and Non-Hispanic American Indian and Alaska Native persons cannot be displayed
  - Kansas – data for Hispanic and Non-Hispanic American Indian and Alaska Native, Asian and Pacific Islander, and Black persons cannot be displayed
  - North Dakota and Wisconsin – Hispanic ethnicity data alone or in combination with any race category cannot be displayed
- For more information, please refer to the *Race recode (W, B, AI, API*) and *Origin Recode NHIA (Hispanic, Non-Hisp*) variable descriptions in this document.

Section: Demographic Data Items
Sex

Sex

SAS Alternate Name	Length	Source of Standard
I220_Sex	1	NAACCR Item #220

Description

Code for the sex of the patient.

Codes

1 Male
2 Female

Considerations for Use

- To get the correct population denominator, select “female” 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.

Section: Demographic Data Items
Age at Diagnosis

Age at Diagnosis

SAS Alternate Name	Length	Source of Standard
I230_AgeDx	3	NAACCR Item #230

Description

Age of the patient at diagnosis in complete years.

Codes

3-character value

Considerations for Use

Population data are not available for this variable; therefore, rates cannot be calculated using this variable. When calculating rates, use the *Age Recode* variable. Age at diagnosis in complete years is used for special analysis, such as modeling.

Age > 99 records are suppressed (have been recoded to 99+).

Section: Demographic Data Items
Age Recode

Age Recode

SAS Alternate Name	Length	Source of Standard
AGEREC	2	Derived based upon NAACCR Item #230

Description

A standard grouping of age at diagnosis into 19 categories.

For further information on creating this variable, see the SAS statements in Appendix I.

Codes

00 = 00 years
01 = 01-04 years
02 = 05-09 years
03 = 10-14 years
04 = 15-19 years
05 = 20-24 years
06 = 25-29 years
07 = 30-34 years
08 = 35-39 years
09 = 40-44 years
10 = 45-49 years
11 = 50-54 years
12 = 55-59 years
13 = 60-64 years
14 = 65-69 years
15 = 70-74 years
16 = 75-79 years
17 = 80-84 years
18 = 85+ years

Considerations for Use

None noted.

Section: Demographic Data Items
Birth Date

Birth Date

SAS Alternate Name	Length	Source of Standard
I240_DOB	4	Derived based upon NAACCR Item #240

Description

Year of birth of the patient.

Codes

4-character date format: YYYY

Considerations for Use

The month and day of birth are not provided for confidentiality reasons.

Age > 99 records are suppressed(have been recoded to 9999).

Section: Demographic Data Items
Economic Status

Economic Status

SAS Alternate Name	Source of Standard
Econ_Status	Derived based upon NAACCR Item #89

Description

County level economic status variable as assigned by the Appalachian Regional Commission. This data item is not available for Puerto Rico.

Codes

See Appendix II for variable coding.

Considerations for Use

- **Exclude Minnesota and Kansas when using this variable**. Caution should also be used with states that have missing county codes. We recommend running a frequency by states to find the number of missing counties (999).
- **Distressed Designation and County Economic Status Classification System**

  The Appalachian Regional Commission (ARC) uses an index-based county economic classification system to identify and monitor the economic status of Appalachian counties. The system involves the creation of a national index of county economic status through a comparison of each county's averages for three economic indicators—three-year average unemployment rate, per capita market income, and poverty rate—with national averages. The resulting values are summed and averaged to create a composite index value for each county. Each county in the nation is then ranked, based on its composite index value, with higher values indicating higher levels of distress.
- ***County Economic Levels***

  Each county is classified into one of five economic status designations, based on its position in the national ranking.

  1. **Distressed**

     Distressed counties are the most economically depressed counties. They rank in the worst 10 percent of the nation's counties.

  2. **At-Risk**

     At-Risk counties are those at risk of becoming economically distressed. They rank between the worst 10 percent and 25 percent of the nation's counties.

  3. **Transitional**

     Transitional counties are those transitioning between strong and weak economies. They make up the largest economic status designation. Transitional counties rank between the worst 25 percent and the best 25 percent of the nation's counties.

4. **Competitive**

   Competitive counties are those that are able to compete in the national economy but are not in the highest 10 percent of the nation's counties. Counties ranking between the best 10 percent and 25 percent of the nation's counties are classified competitive.

5. **Attainment**

   Attainment counties are the economically strongest counties. Counties ranking in the best 10 percent of the nation's counties are classified attainment.

- A description of the source and methodology of the Appalachian Regional Commission is available at:

  Classifying Economic Distress in Appalachian Counties - Appalachian Regional Commission (arc.gov)

Section: Demographic Data Items
Rural-urban Continuum 2013

Rural-urban Continuum 2013

SAS Alternate Name	Length	Source of Standard
Ruralurban_continuum_2013	2	NAACCR Item #3312

Description

The *RuralUrban Continuum (2013)* codes (usually referred to as the Beale Codes) separate counties into four metropolitan and six non-metropolitan categories, based on the size their populations and form a classification scheme that distinguishes metropolitan counties by size and nonmetropolitan counties by degree of urbanization and proximity to metro areas.

Codes

See Appendix II for coding

Considerations for Use

- These codes are derived electronically by the central cancer registry using patients' county at diagnosis. In instances where the central cancer registry is unable to submit county information, the codes are derived and submitted by the registry. FIPS state and county code mappings to Beale Codes can be obtained in an Excel file at http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx
- The code is a 9-point continuum, transmitted in standard NAACCR record form with a leading 0, (01-09). Abstractors do not enter these codes.
- Areas that are not included in the *Rural-Urban Continuum* code table, such as U.S. territories (other than Puerto Rico) are coded 98. Records where the If County at DX = 999, the *Rural-Urban Continuum* are coded 99.

Code Number	Description
Metropolitan Counties (01-03)	
01	Counties in metro areas of 1 million population or more
02	Counties in metro areas of 250,000 to 1 million population
03	Counties in metro areas of fewer than 250,000 population

Code Number	Description
Nonmetropolitan Counties (04-09)	
04	Urban population of 20,000 or more, adjacent to a metro area
05	Urban population of 20,000 or more, not adjacent to a metro area
06	Urban population of 2,500 to 19,999, adjacent to a metro area
07	Urban population of 2,500 to 19,999, not adjacent to a metro area
08	Completely rural or less than 2,500 urban population, adjacent to a metro area
09	Completely rural or less than 2,500 urban population, not adjacent to a metro area
98	Program run, but: (1) area is not included in Rural-Urban Continuum code table, or (2) record is for resident outside of state of reporting institution
99	Unknown
Blank	Program not run; record not coded

Section: Cancer Identification Data Items
Sequence Number – Central

Sequence Number – Central

SAS Alternate Name	Length	Source of Standard
I380_SeqNoCntrl	2	NAACCR Item #380

Description

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

Codes

In Situ/Malignant as Federally Required based on Diagnosis Year:

00 One primary in the patient's lifetime
01 First of two or more primaries
02 Second of two or more primaries
..
..
59 Fifty-ninth or higher of fifty-nine or more primaries
99 Unspecified or unknown sequence number of federally required *in situ* or malignant tumors. Sequence number 99 can be used if there is a malignant tumor and its sequence number is unknown. If there is known to be more than one malignant tumor, then the tumors must be sequenced.

Non-malignant Tumor as Federally Required based on Diagnosis Year or State/Province Defined:

60 One non-malignant tumor or central registry-defined neoplasm
61 First of two or more non-malignant tumor or central registry-defined neoplasms
62 Second of two or more non-malignant tumor or central registry-defined neoplasms
.. ..
88 Unspecified or unknown sequence number for non-malignant tumor or central registry-defined neoplasms. (Sequence number 88 can be used if there is a non-malignant tumor and its sequence number is unknown. If there is known to be more than one non-malignant tumor, then the tumors must be sequenced.)
98 Cervix carcinoma *in situ* (CIS)/CIN III, Diagnosis Years 1996-2002.

Table showing the sequence number series to use by type of neoplasm.

Neoplasm	SeqNum-Central
***In Situ*/Malignant as Federally Required based on Diagnosis Year**	**(Numeric Series)**
In Situ (behavior code = 2) (Cervix CIS/CIN III, Diagnosis Year before 1996) (includes VIN III, VAIN III, AIN III)	00 -- 59
Malignant (behavior code = 3)	00 -- 59

Juvenile Astrocytoma, Diagnosis Year 2001+ (*)	00 -- 59
Invasive following *In Situ*--New primary as defined by CoC	00 -- 59
Invasive following *In Situ*--New primary as defined by SEER	00 -- 59
Unspecified Federally Required Sequence Number or Unknown	99
Non-malignant Tumor as Federally Required based on Diagnosis Year or State/Province Registry-Defined	
Examples:	
Non-malignant Tumor/Benign Brain	60 -- 87
Borderline Ovarian, Diagnosis Year 2001+	60 -- 87
Other Borderline/Benign	60 -- 87
Skin SCC/BCC	60 -- 87
PIN III	60 -- 87
Cervix CIS/CIN III, Diagnosis Year 2003+	60 -- 87
Unspecified Non-malignant Tumor or Central Registry-Defined Sequence Number	88
Cervix CIS/CIN III, Diagnosis Year 1996-2002	98

*Juvenile astrocytomas should be reported as 9421/3.
Note: Conversion Guidance: The sequence numbers for neoplasms whose histologies were associated with behavior codes that changed from *in situ*/malignant to benign/borderline or vice versa during the conversion from ICD-O-2 to ICD-O-3 should not be re-sequenced.

Considerations for Use

- The sequence number may change over the patient's lifetime. If the patient was diagnosed with a single reportable neoplasm, and later diagnosed with a second reportable neoplasm, the sequence code for the first neoplasm changes from 00 to 01. A central registry may find that a patient with one or more known neoplasms had an earlier reportable neoplasm that had been unknown to the registry. Typically, a re-evaluation of all related sequence numbers is required whenever an additional neoplasm is identified.
- Standards define which neoplasms are reportable. It is assumed that these standards are the minimum definition of reportability. Individual central cancer registries may define additional neoplasms as reportable. Variability of assigning sequence numbers over time may exist for different registries, which may impact the coding of this variable.
- Because the time period of *Sequence Number* is a person's lifetime, reportable neoplasms not included in the central registry (those that occur outside the registry catchment area or before the reference date) also are allotted a sequence number. For example, a registry may contain a single record for a patient with a sequence number of 02 because the first reportable neoplasm preceded the central registry's reference date.
- If two or more reportable neoplasms are diagnosed at the same time, the lowest sequence number is assigned to the diagnosis with the worst prognosis. If no difference in prognosis is evident, the decision is arbitrary.

- Reportable non-malignant tumors diagnosed on or after January 1, 2004 are represented by sequence numbers labeled as "…state registry-defined neoplasm". Timing rules for sequencing these neoplasms are the same as timing rules for sequencing required *in situ* or invasive neoplasms.
- The 2007 Multiple Primary and Histology Coding Rules and 2023 Solid Tumor Rules may also affect the sequence number.

Section: Cancer Identification Data Items
Year of diagnosis

Year of diagnosis

SAS Alternate Name	Length	Source of Standard
DXYEAR	4	Derived based on NAACCR Item #390

Description

Year of diagnosis by a recognized medical practitioner for the cancer being reported whether clinically or microscopically confirmed. This dataset contains records with a diagnosis year of 1998-2020.

Puerto Rico data are available for 2005-2020 diagnosis years only.

Codes

4-character date format: YYYY

Considerations for Use

YYYY – when year is known and valid
Blank – when no known date applies

Section: Cancer Identification Data Items
Date of Diagnosis

Date of Diagnosis

SAS Alternate Name	Length	Source of Standard
I390_DateDx	6	Derived based upon NAACCR Item #390

Description

Date of initial diagnosis by a recognized medical practitioner for the cancer being reported whether clinically or microscopically confirmed. This dataset contains records with a diagnosis year of 1998-2020.

Puerto Rico data are available for 2005-2020 diagnosis years only.

Codes

6-character date format: YYYYMM

Considerations for Use

- The day of diagnosis is not provided for confidentiality reasons.
- Only valid portions of the date are included in this dataset.
- Below are the common formats to handle the situation where only certain components of date are known–
  YYYYMM – when year and month are known and valid
  YYYY – when year is known and valid, and month is unknown or invalid
  Blank – when no known date applies

Section: Cancer Identification Data Items
Primary Site

Primary Site

SAS Alternate Name	Length	Source of Standard
Primary_Site	3	NAACCR Item #400

Description

Code for the *primary site* of the tumor being reported using ICD-O-3.

Codes

See Appendix II for coding

Considerations for Use

See ICD-O-3 Topography Section for the codes for *primary site*. Reference: Fritz A et al., editors. International Classification of Diseases for Oncology. Third Edition. Geneva: World Health Organization; 2000.

Consider reviewing the variables "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/.

Section: Cancer Identification Data Items
Laterality

Laterality

SAS Alternate Name	Length	Source of Standard
I410_Laterality	1	NAACCR Item #410

Description

Code for 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.

Starting with cases diagnosed January 1, 2004, and later, laterality is coded for select invasive, benign, and borderline primary intracranial and CNS tumors.

Codes

0 Not a paired site
1 Right: origin of primary
2 Left: origin of primary
3 Only one side involved, right or left origin unspecified
4 Bilateral involvement at time of diagnosis, lateral origin unknown for a single primary; or both ovaries involved simultaneously, single histology; bilateral retinoblastomas; bilateral Wilms tumors
5 Paired site: midline tumor (effective with 1/1/2010 dx)
9 Paired site, but no information concerning laterality

Section: Cancer Identification Data Items
Grade

Grade

SAS Alternate Name	Length	Source of Standard
I440_Grade	1	NAACCR Item #440

Description

Code for the grade or degree of differentiation of the reportable tumor. For lymphomas and leukemias, field also is used to indicate T-, B-, Null-, or NK-cell origin.

Codes

Histologic Grading and Differentiation

1	Grade I	Well differentiated Differentiated, NOS
2	Grade II	Moderately differentiated Moderately well differentiated Intermediate differentiation
3	Grade III	Poorly differentiated
4	Grade IV	Undifferentiated Anaplastic

Immunophenotype Designation for Lymphomas and Leukemias

5 T-cell
6 B-cell
7 Null cell
8 NK (natural killer) cell

Comment: Use the most recent hematopoietic and lymphoid rules for assigning grades 5-8.

9 Grade/differentiation unknown, not stated, or not applicable

Considerations for Use

This variable only available through dx year 2017.

Reference: Fritz A et al., editors. International Classification of Diseases for Oncology. Third Edition. Geneva: World Health Organization; 2000.

The practice of grading varies greatly among pathologists throughout the world, and many malignant tumors are not routinely graded. 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 STORE manual (https://www.facs.org/quality-programs/cancer/ncdb/call-for-data/cocmanuals). Each module has an abstracting, coding, and staging section, which has a morphology and grading sub-section. Some modules, but not all, contain notes about the grading system that may have been used to code grade. Currently, this dataset does not contain a 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 artificial increase in higher grade from 2002 to 2003. Additional review showed that the International Society of Urologic Pathologists (ISUP) in conjunction with the WHO 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. More information about grade migration is available:

1. Luthringer DJ, Gross M. Gleason Grade Migration: Changes in Prostate Cancer Grade in the Contemporary Era. *PCRI Insights* 2001; 9(3).
2. Jani AB, Master VA, Rossi PJ, Liauw SL, Johnstone PAS. Grade migration in prostate cancer: an analysis using the Surveillance, Epidemiology, and End Results registry. *Prostate Cancer and Prostatic Diseases* 2007; 10: 347–351.
3. Thompson IM, Canby-Hagino E, Lucia MS. Stage Migration and Grade Inflation in Prostate Cancer: Will Rogers Meets Garrison Keillor. *Journal of the National Cancer Institute* 2005; 97(17): 1236-7.

Section: Cancer Identification Data Items
Grade Clinical

Grade Clinical

SAS Alternate Name	Length	Source of Standard
I3843_GradeClinical	1	NAACCR Item #3843

Description

This data item records the grade of a solid primary tumor before any treatment (surgical resection or initiation of any treatment including neoadjuvant).
For cases diagnosed January 1, 2018, and later, this data item, along with *Grade Pathological* and Grade Post-Neoadjuvant, replaces the data item *Grade* as well as site specific factors (SSF's) for cancer sites with alternative grading systems (e.g., breast [Bloom-Richardson], prostate [Gleason]).

Codes

See Appendix II for coding

Considerations for Use

- Data are available for cases diagnosed after 2018. This data item was not collected for cases diagnosed 2001–2017.
- For cases that are eligible for American Joint Committee on Cancer (AJCC) staging, the recommended grading system is specified in the AJCC Cancer Staging Manual chapter. The AJCC chapter-specific grading systems (codes 1-5) take priority over the generic grade definitions (codes A-E, L, H, 9). For cases that are not eligible for AJCC staging, if the recommended grading system is not documented, the generic grade definitions apply.
- Refer to the Site Specific Data Item (SSDI) Manual and Grade manual that corresponds with the year the cases were diagnosed for additional site-specific instructions.

Section: Cancer Identification Data Items
Grade Pathological

Grade Pathological

SAS Alternate Name	Length	Source of Standard
I3844_GradePathological	1	NAACCR Item #3844

Description

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

For cases diagnosed January 1, 2018, and later, this data item, along with Grade Clinical and Grade Post-Neoadjuvant, replaces the data item *Grade* as well as site specific factors (SSF's) for cancer sites with alternative grading systems (e.g., breast [Bloom-Richardson], prostate [Gleason]).

Codes

See Appendix II for coding

Considerations for Use

- Data are available for cases diagnosed after 2018. This data item was not collected for cases diagnosed from 2001–2017.
- For cases that are eligible for American Joint Committee on Cancer (AJCC) staging, the recommended grading system is specified in the AJCC Cancer Staging Manual chapter. The AJCC chapter-specific grading systems (codes 1-5) take priority over the generic grade definitions (codes A-E, L, H, 9). For cases that are not eligible for AJCC staging, if the recommended grading system is not documented, the generic grade definitions apply.
- Refer to the Site Specific Data Item (SSDI) Manual and Grade manual for additional site-specific instructions.

Section: Cancer Identification Data Items
Grade Post Therapy Path

Grade Post Therapy Path

SAS Alternate Name	Length	Source of Standard
I3845_GradePostTherapy	1	NAACCR Item #3845

Description

This data item records the grade of a solid primary tumor that has been resected following neoadjuvant therapy. The highest grade documented from the surgical treatment resection specimen of the primary site following neoadjuvant therapy will be recorded. For cases diagnosed January 1, 2018 and later, this data item, along with Grade Clinical, Grade Pathological, and Grade Post Therapy Clin (yc), replaces all previous grade related data items, including NAACCR Data Item *Grade (#440)* and Collaborative Stage Site-Specific Factors (SSF's) (2004-2017) for cancer sites with alternative grading systems (e.g., breast [Bloom-Richardson], prostate [Gleason]).

Codes

See Appendix II for coding

Considerations for Use

- Data are available for cases diagnosed after 2018. This data item was not collected for cases diagnosed from 2001–2017.
- For cases that are eligible for American Joint Committee on Cancer (AJCC) staging, the recommended grading system is specified in the AJCC Cancer Staging Manual chapter. The AJCC Chapter-specific grading systems (codes 1-5, H, L, M, S and 9) take priority over the generic grade definitions (codes A-E). For those cases that are not eligible for AJCC staging, if the recommended grading system is not documented, the generic grade definitions may apply.
- Refer to the Site Specific Data Item (SSDI) Manual and Grade manual for additional site-specific instructions.

Section: Cancer Identification Data Items
Diagnostic Confirmation

Diagnostic Confirmation

SAS Alternate Name	Length	Source of Standard
I490_DxConf	1	NAACCR Item #490

Description

Code for the best method of diagnostic confirmation of the cancer being reported at any time in the patient's history.

Codes

1 Positive histology
2 Positive cytology
3 Positive histology PLUS – positive immunophenotyping AND/OR positive genetic studies (Used only for hematopoietic and lymphoid neoplasms M-9590/3-9992/3)
4 Positive microscopic confirmation, method not specified
5 Positive laboratory test/marker study
6 Direct visualization without microscopic confirmation
7 Radiography and/or other imaging techniques without microscopic confirmation
8 Clinical diagnosis only (other than 5, 6, or 7)
9 Unknown whether or not microscopically confirmed; death certificate only

Note: Code 3 (used only for hematopoietic and lymphoid neoplasms M-9590/3-9992/3) was adopted for use effective with 2010 diagnoses.

Considerations for Use

None noted.

Section: Cancer Identification Data Items
Type of Reporting Source

Type of Reporting Source

SAS Alternate Name	Length	Source of Standard
I500_TypeRptSrc	1	NAACCR Item #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 this item 4).

The code in this field can be used to explain why information may be incomplete on a tumor. For example, death certificate only cases have unknown values for many data items, so one may want to exclude them from some analyses. The field also is used to monitor the success of non-hospital case reporting and follow-back mechanisms. All population-based registries should have some death certificate-only cases where no hospital admission was involved, but too high a percentage can imply both shortcomings in case-finding and that follow-back to uncover missed hospital reports was not complete.

Codes

1 Hospital inpatient; Managed health plans with comprehensive, unified medical records
2 Radiation Treatment Centers or Medical Oncology Centers (hospital-affiliated or independent)
3 Laboratory only (hospital-affiliated or independent)
4 Physician's office/private medical practitioner (LMD)
5 Nursing/convalescent home/hospice
6 Autopsy only
7 Death certificate only
8 Other hospital outpatient units/surgery centers

Considerations for Use

Codes are assigned in the following priority order: 1, 2, 8, 4, 3, 5, 6, 7. This prioritizes laboratory reports over nursing home reports. The source facilities included in the code 1 (hospital inpatient and outpatient) were split in 2006 between codes 1, 2, and 8. Sources coded with '8' would include, but would not be limited to, outpatient surgery and nuclear medicine services.

Section: Cancer Identification Data Items
Histologic Type ICD-O-3

Histologic Type ICD-O-3

SAS Alternate Name	Length	Source of Standard
I522_HistTypeICDO3	4	NAACCR Item #522

Description

Codes for the histologic type of the tumor being reported using ICD-O-3. ICD-O-3 was adopted as the standard coding system for tumors diagnosed in 2001 and later. Tumors diagnosed prior to 2001 have been converted from ICD-O-2. Effective with cases diagnosed in 2010 and forward, this item also includes codes for new terms as per the 2008 WHO Hematopoietic/Lymphoid publication.

Codes

See NAACCR data dictionary for coding: https://www.naaccr.org/data-standards-data-dictionary/

Considerations for Use

Reference: Fritz A et al., editors. International Classification of Diseases for Oncology. Third Edition. Geneva: World Health Organization; 2000.

See ICD-O-3, Morphology Section and the SEER Hematopoietic database (http://seer.cancer.gov/tools/heme/).

Section: Cancer Identification Data Items
Behavior Code ICD-O-3

Behavior Code ICD-O-3

SAS Alternate Name	Length	Source of Standard
I523_BehavICDO3	1	NAACCR Item #523

Description

Code for the behavior of the tumor being reported using ICD-O-3. ICD-O-3 was adopted as the standard coding system for tumors diagnosed in 2001 and later. Tumors diagnosed prior to 2001 have been converted from ICD-O-2.

Juvenile astrocytoma is coded as borderline in ICD-O-3; North American registries report as 9421/3.

Codes

0 Benign
1 Uncertain whether benign or malignant
Borderline malignancy
Low malignant potential
Uncertain malignant potential
2 Carcinoma in situ
Intraepithelial
Noninfiltrating
Noninvasive
3 Malignant, primary site

Considerations for Use

Reference: Fritz A et al., editors. International Classification of Diseases for Oncology. Third Edition. Geneva: World Health Organization; 2000.

Section: Cancer Identification Data Items
Behavior Recode for Analysis Derived/WHO2008

Behavior recode for analysis derived/WHO2008

SAS Alternate Name	Source of Standard
Behavior_Recode_For_Analysis	Derived based upon NAACCR Items #400, #522, and #523

Description

The purpose of this variable is to allow for selection of behavior codes that are consistent between ICD-O-2 and ICD-O-3. ICD-O-3 is used to code cases diagnosed on or after January 1, 2001. Codes that are newly malignant in ICD-O-3 and codes that are no longer malignant in ICD-O-3 (e.g., borderline ovarian cancers) show up as invalid.

Codes

See Appendix II for coding.

Considerations for Use

See Appendix 6 in ICD-O-3 for a list of histologies that changed behavior. Reference: Fritz A et al., editors. International Classification of Diseases for Oncology. Third Edition. Geneva: World Health Organization; 2000.

Section: Cancer Identification Data Items
Site Recode ICD-O-3/WHO 2008

Site recode ICD-O-3/WHO 2008

SAS Alternate Name	Source of Standard
Site_Recode_ICD_O_3_WHO_2008	Derived based upon NAACCR Items #400 and #522

Description

The values of the primary site recode variable are based on the primary site and histology data fields submitted by the registries. The site recode variables define the major cancer sites that are commonly used in the reporting of cancer incidence data. This recode is defined by the SEER program.

Codes

See Appendix II for coding.

Considerations for Use

Reference for Primary Site Recode for ICD-O-3 is:
http://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html

Section: Cancer Identification Data Items
AYA Site Recode 2020 Revision

AYA Site Recode 2020 Revision

SAS Alternate Name	Source of Standard
AYA_site_recode_2020_revision	Derived from NAACCR Items #400, #522, and #523

Description

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

Codes

See Appendix II for coding

Considerations for Use

- This recode variable is defined by the SEER Program and is based on the classification scheme proposed by RD Barr and colleagues. Refer to the AYA Site Recode for the full list of 318 groups and additional information.
- More information is available at https://seer.cancer.gov/ayarecode/aya-2020.html

Section: Cancer Identification Data Items
Lymphoid Neoplasm Recode 2021 Revision

Lymphoid neoplasm recode 2021 revision

2021SAS Alternate Name	Source of Standard
Lymphoid_neoplasm_recode_2021	Derived based upon NAACCR Items #400, #522 and #523

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.

Codes

See Appendix II for coding.

Considerations for Use

- This recode variable is defined by the SEER program. It was adapted from a proposed nested classification of lymphoid neoplasms in: Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph) (2007).
- 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/lymphoma-2021.html

Section: Cancer Identification Data Items
SEER-Modified International Classification of Childhood Cancer (ICCC) Recode

ICCC site recode extended ICD-O-3/WHO 2008

SAS Alternate Name	Source of Standard
ICCCSiteRecExtendedICDO3WHO2008	Derived based upon NAACCR Items #400, #522, and #523

Description

The values of *the SEER-modified International Classification of Childhood Cancer recode* variable are based on the primary site and histology data fields submitted by the registries. The classification of childhood cancer is based on tumor morphology rather than, as for adults, the site of the tumor. These recodes were adapted by the SEER program from groupings developed by the World Health Organization.

Codes

See Appendix II for coding.

Considerations for Use

Note that beginning with data released in 2006, the grouping of childhood cancers is based on ICD-O-3 instead of ICD-O-2.

Reference for the ICCC recodes is:
http://seer.cancer.gov/iccc/

Section: Stage/Prognostic Factors Data Items
SEER Summary Stage 2000

SEER Summary Stage 2000

SAS Alternate Name	Length	Source of Standard
I759_SS2000	1	NAACCR Item #759

Description

Code for the summary stage at the initial diagnosis or treatment of the reportable tumor. For site-specific definitions of categories, see SEER *Summary Staging Manual 2000*.

Summary stage should include all information available through completion of surgery(ies) in the first course of treatment or within 4 months of diagnosis in the absence of disease progression, whichever is longer.

Codes

0 *In situ*
1 Localized
2 Regional, direct extension only
3 Regional, regional lymph nodes only
4 Regional, direct extension and regional lymph nodes
5 Regional, NOS
7 Distant
8 Not applicable
9 Unstaged

Note: Code 8 was added effective with cases diagnosed in 2004 and forward to be used when there is not an applicable code to reflect stage (e.g., benign brain, borderline ovarian).

Note: See also the item *Derived SS2000 [3020]* for the value of *SEER Summary Stage 2000* as generated by the collaborative Staging algorithm.

Considerations for Use

Summary stage is a required variable. The correct data item to use (and corresponding code manual) is determined by the year in which the cancer was diagnosed. *SEER Summary Stage 2000* is used for tumors diagnosed 2001-2003. NPCR funded central cancer registries also used *SEER Summary Stage 2000* for diagnosis year 2016 and 2017.

See Cautionary Notes – Stage for additional information.

Section: Stage/Prognostic Factors Data Items
SEER Summary Stage 1977

SEER Summary Stage 1977

SAS Alternate Name	Length	Source of Standard
I760_SS1977	1	NAACCR Item #760

Description

Code for summary stage at the initial diagnosis or treatment of the reportable tumor. This has traditionally been used by central registries to monitor time trends. For site-specific definitions of categories, see the SEER Summary Staging Guide.

SEER Summary Stage 1977 is limited to information available within 2 months of the date of diagnosis. NAACCR approved extension of this time period to 4 months for prostate tumors diagnosed beginning January 1, 1995.

Codes

0 In situ
1 Localized
2 Regional, direct extension only
3 Regional, regional lymph nodes only
4 Regional, direct extension and regional lymph nodes
5 Regional, NOS
7 Distant
8 Not applicable
9 Unstaged

Note: Code 8 was added effective with cases diagnosed in 2004 and forward to be used when there is not an applicable code to reflect stage (e.g., benign brain, borderline ovarian).

Considerations for Use

Summary stage is a required variable. The correct data item to use (and corresponding code manual) is determined by the year in which the cancer was diagnosed. *SEER Summary Stage 1977* is used for tumors diagnosed between 1998-2000.

See Cautionary Notes – Stage for additional information.

Section: Stage/Prognostic Factors Data Items
Derived SEER Summary State 2000

Derived SS2000

SAS Alternate Name	Length	Source of Standard
I3020_DerivedSS2000	1	NAACCR Item #3020; AJCC

Description

This item is the "*SEER Summary Stage 2000*" derived from the CS algorithm effective with 2004 diagnosis year.

The Collaborative Stage Data Collection System was designed by a joint task force including representatives from SEER, ACoS, CDC, NAACCR, NCRA, CCCR, CPAC, and AJCC, to provide a single uniform set of codes and rules for coding stage information to meet the needs of all of the participating standard setters. When CS data items are coded, a computer algorithm provides the derivation of *SEER Summary Stage 2000*.

Codes

1-character code

See NAACCR data dictionary for details: https://www.naaccr.org/data-standards-data-dictionary/

Considerations for Use

Refer to the Collaborative Stage Data Collection System Manual and Coding Instructions (http://cancerstaging.org/cstage/Pages/default.aspx) for rules and site-specific codes and coding structures.

Records in this dataset should have a Derived SS2000 value for diagnosis years 2004-2017 for SEER only registries and 2004-2015 for NPCR registries. This data item is usually blank for records in this dataset with a diagnosis year prior to 2004 (1998-2003).

The data item SEER Summary Stage 1977 provides stage information for records with a diagnosis year of 1998-2000 and SEER Summary Stage 2000 provides stage information for records with a diagnosis year of 2001-2003.

To study historical trends in stage, Summary Stage should be selected according to the following table:

Diagnosis Years	Summary Stage Version
1998-2000	Summary Stage 1977
2001-2003	Summary Stage 2000
2004-2015	Derived Summary Stage 2000 (see note above)
2016-2017	Summary Stage 2000 or Derived Summary Stage 2000 (see note above)

Previous data quality analyses identified concerns with the information reported in this variable, such as conflicts between the coded CS Extension and Behavior variables; e.g. in situ behavior with an extension indicating an invasive lesion. It is felt that subsequent training and

implementation of additional electronic data edits have greatly improved the validity and reliability of the staging information. If there are concerns about stage distributions resulting from data queries, please contact CDC (uscsdata@cdc.gov).

See Cautionary Notes – Stage for additional information.

Section: Stage/Prognostic Factors Data Items
Summary Stage 2018

Summary Stage 2018

SAS Alternate Name	Length	Source of Standard
I764_SS2018	1	NAACCR Item #764

Description

Code for summary stage at the initial diagnosis or treatment of the reportable tumor. *Summary Stage 2018* is used for tumors diagnosed January 1, 2018 and later.

Codes

See Appendix II for coding

Considerations for Use

The 2018 version of Summary Stage applies to every site and/or histology combination, including lymphomas and leukemias. Summary Stage uses all information available in the medical record; in other words, it is a combination of the most precise clinical and pathological documentation of the extent of disease. More information is available at https://seer.cancer.gov/tools/ssm/.

Section: Stage/Prognostic Factors Data Items
Merged Summary Stage

Merged Summary Stage

SAS Alternate Name	Length	Source of Standard
Merged_Summary_Stage	1	NPCR, combined from NAACCR Items #759, #760, #764, and #3020

Description

This is a merged stage variable created using four other variables: *SEER Summary Stage 1977*, *SEER Summary Stage 2000*, *Summary Stage 2018,* and *Derived SS2000*.

Codes

See the Appendix II for coding.

Considerations for use

The coding logic for this merged variable is:

- If a case was diagnosed between 1998 and 2000, stage at diagnosis is recorded using the *SEER Summary Stage 1977* variable value.
- 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 and it was reported to an NPCR registry, then the stage at diagnosis is recorded using the *Derived SEER Summary Stage 2000* variable value.
- If a case was diagnosed between 2016 and 2017 and it was reported to an NPCR registry, then the stage at diagnosis is recorded using the SEER *Summary Stage 2000*.
- If a case was diagnosed between 2004 and 2017 and it was reported by a SEER registry (Connecticut, Hawaii, Iowa, or New Mexico), 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 2000* was blank, but *SEER Summary Stage* 2000 had a value of *local*, then the merged variable was coded as local stage. Otherwise, the merged variable is left blank for that record.
- If a case was diagnosed after 2018, stage at diagnosis is recorded using the *Summary Stage 2018* variable value.

See Cautionary Notes – Stage for additional information.

Section: Stage/Prognostic Factors Data Items
CS Site Specific Factor 1

CS Site Specific Factor 1

SAS Alternate Name	Length	Source of Standard
I2880_CSSSF1	3	NAACCR Item #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 for:

- **Brain and Other Nervous System sites: World Health Organization (WHO) Grade Classification**.

Codes

See the Appendix II for coding.

Considerations for Use

- Data for this variable are available for:
  - **Brain and Other Nervous System** starting with **diagnosis years 2011–2017**.
  - **CSSSF1 does not include sites C300 and C301**
- For the site-specific codes, please refer to the Collaborative Stage Data Collection System
  - **Brain:** World Health Organization (WHO) Grade Classification
  - **Other parts of central nervous system:** World Health Organization (WHO) Grade Classification

Section: Stage/Prognostic Factors Data Items
Merged Estrogen Receptor

Merged Estrogen Receptor

SAS Alternate Name	Source of Standard
Merged_estrogen_receptor	Derived from NAACCR Items #2880 and #3827

Description

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

Codes

See Appendix II for coding

Considerations for Use

- Data for this variable are available for female breast cancer cases diagnosed ≥2004.
- For more information –
    - Collaborative Stage Data Collection System for the specific codes for CS SSF1, Breast Estrogen Receptor Assay, available at http://web2.facs.org/cstage0205/breast/Breast_jag.html.
    - Site Specific Data Items (SSDI), NAACCR Cancer Schema List

Section: Stage/Prognostic Factors Data Items
Merged Progesterone Receptor

Merged Progesterone Receptor

SAS Alternate Name	Source of Standard
Merged_progesterone_receptor	Derived from NAACCR Items #2890 and #3915

Description

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

Codes

See Appendix II for coding

Considerations for Use

- Data for this variable are available for female breast cancer cases diagnosed ≥2004.
- For more information –
    - 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.
    - Site Specific Data Items (SSDI), NAACCR Cancer Schema List

Section: Stage/Prognostic Factors Data Items
Merged HER2 Summary

Merged HER2 Summary

SAS Alternate Name	Source of Standard
Merged_HER2_summary	Derived from NAACCR Items #2869 and #3855

Description

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

Codes

See Appendix II for coding

Considerations for Use

- Data for this variable are available for female breast cancer cases diagnosed ≥2010.
- For more information –
    - Collaborative Stage Data Collection System for the specific codes for CS SSF15, Breast HER2 Summary Result of Testing, available at http://web2.facs.org/cstage0205/breast/Breast_sbg.html.
    - Site Specific Data Items (SSDI), NAACCR Cancer Schema List

Section: Treatment—First Course
RX Summ—Surgery Primary Site

RX Summ—Surgery Primary Site

SAS Alternate Name	Length	Source of Standard
I1290_RxSummSurgPrimSite	2	NAACCR Item #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.

Codes

00	None
10-19	Site-specific code; tumor destruction
20-80	Site-specific codes; resection
90	Surgery, NOS
98	Site specific codes; special
99	Unknown

Considerations for use

- Data for this variable are available starting with **diagnosis year 2003**.
- In addition to the site-specific codes, refer to the most recent version of *STORE* and SEER Program Code manual for additional instructions:
  - STORE manual - https://www.facs.org/quality-programs/cancer/ncdb/call-for-data/cocmanuals
  - SEER Program Code manual - https://seer.cancer.gov/tools/codingmanuals

Section: Treatment—First Course
Merged Radiation

Merged Radiation

SAS Alternate Name	Source of Standard
Merged_Radiation	Derived from NAACCR Items #1360, #1506 and #1570

Description

This is a user-defined variable created for this database that merges *RX SUMM–Radiation* (NAACCR item 1360), *Phase I Radiation Treatment Modality* (NAACCR item 1506), and *Rad–Regional RX Modality* (NAACCR item 1570) and provides treatment information.

Codes

See Appendix II for recoding.

Considerations for use

- This variable is only available for female breast, colorectal and for cases submitted by NPCR central cancer registries.
- Data for this variable are available starting with diagnosis year 2010.

Section: Over-ride Flags Data Items
Over-ride Age/Site/Morph

Over-ride Age/Site/Histology Inter-field Review (Inter-field Edit 15)

SAS Alternate Name	Length	Source of Standard
I1990_ORAgeSiteMorph	1	NAACCR Item #1990; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Age, Primary Site, Morphology ICDO3 (SEER IF15)	Identifies records with an unusual occurrence of a particular age/site/histology combination for a given age group
Age, Primary Site, Morph ICDO3--Adult (SEER)	Identifies records with an unusual occurrence of a particular age/site/histology combination for a given age group in records with an age at diagnosis ≥15
Age, Primary Site, Morph ICDO3--Pediatric (NPCR)	Identifies records with an unusual occurrence of a particular age/site/histology combination for a given age group in records with an age at diagnosis 00-14

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

Some cancers occur almost exclusively in certain age groups.

Edits of the Age, Primary Site, and Morphology require review if a site/morphology combination occurs in an age group for which it is extremely rare. The edit Age, Primary Site, Morph ICDO3--Adult (SEER) edits ca7ses with an Age at Diagnosis of 15 and older. The edit Age, Primary Site, Morph ICDO3--Pediatric (NPCR) edits cases with an Age at Diagnosis of less than 15. The edit Age, Primary Site, Morphology ICDO2 (SEER IF15) contains logic for all ages.

Instructions for Coding

1. The data item is to be left blank if the program does not generate an error message (and if the case was not diagnosed in utero) for the edits of the Age, Primary Site, Morphology.

2. Any identified errors should have been corrected for the case if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
3. Code 1 or 3 indicates that a review of data items in the error or warning message confirmed all were correct.

Codes

1 Reviewed and confirmed that age/site/histology combination is correct as reported
2 Reviewed and confirmed that case was diagnosed in utero
3 Reviewed and confirmed that conditions 1 and 2 both apply
Blank Not reviewed or reviewed and corrected.

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in an age, site, morphology combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride SeqNo/DxConf

Over-ride Sequence Number/Diagnostic Confirmation Inter-field Review (Inter-field Edit 23)

SAS Alternate Name	Length	Source of Standard
I2000_ORSeqNoDxConf	1	NAACCR Item #2000; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edit in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Diagnostic Confirm, Seq Num--Central (SEER IF23)	Identifies records with multiple primary cancers where at least one primary cancer is not microscopically confirmed

Rationale

Some edits check for code combinations that are impossible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

- The edit checks if the case is one of multiple primaries and is not microscopically confirmed or has only positive lab test/marker studies (i.e., Diagnostic Confirmation >5) and tumor sequence number >00 (more than one primary).
- The edit is skipped if the Sequence Number--Central is in the range of 60-99.

Instructions for Coding

1. The data item is left blank if the program does not generate an error message for the Diagnostic Confirmation and Sequence Number Central edit.
2. Any identified errors should have been corrected for the case if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
3. Code 1 indicates that a review of data items in the error or warning message verified that there are multiple primary cancers of specific sites in which at least one diagnosis was not microscopically confirmed.

Codes

1 Reviewed and confirmed as reported
Blank Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a diagnostic confirmation and sequence number-central combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Site/Lat/Sequence Number

Over-ride Site/Histology/Laterality/Sequence Number Inter-record Review (Inter-record Edit 09)

SAS Alternate Name	Length	Source of Standard
I2010_ORSiteLatSeqNo	1	NAACCR Item #2010; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

Mature central cancer registries can have up to 15-20% multiple primary data. In order to validate coded values across multiple tumor records for a single patient, inter-record edits are applied to the data. Inter-record edits compare data recorded across more than one record, and are commonly applied across tumor records for a patient that has multiple tumors. These edits compare codes or groups of codes recorded in the same data item(s) between each of the tumor records for the patient. For example, one inter-record edit compares the sequence numbers of multiple tumors for the same patient with their dates of diagnosis to ensure that the sequence numbers have been assigned in the correct chronological order based on diagnosis date.

This over-ride is used with the following Inter-record Edit from the SEER Program:

Inter-record Edit	Description
Verify Same Primary Not Reported Twice for a Person (SEER IR09)	Identifies records with multiple primary cancers where the date of diagnosis and primary cancer site are within a specified range but the sequence number-central is different

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

Verify Same Primary Not Reported Twice for a Person (SEER IR09) applies to paired organs and does not allow two cases with the same primary site group, laterality and three digit histology code. This edit verifies that the same primary is not reported twice for a person.

Instructions for Coding

1. The data item is left blank if the program does not generate an error message for the edit Verify Same Primary Not Reported Twice for a Person (SEER IR09).

2. Any identified errors should have been corrected if the records are determined to be the same primary cancer. The data item should be blank for records where identified errors were corrected.
3. Code 1 indicates that the case was reviewed and verified that the patient had multiple primaries of the same histology (3 digit) in the same primary site group.

Codes

1	Reviewed and confirmed as reported
Blank	Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submission. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a primary site, histology, laterality, and sequence number-central combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Site/Type

Over-ride Site/Type Inter-field Review (Inter-field Edit 25)

SAS Alternate Name	Length	Source of Standard
	1	NAACCR Item #2030; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Primary Site, Morphology-Type, Behavior ICDO3 (SEER IF25)	Identifies records where the site/histology/behavior combination is not in the SEER Site/Histology Validation List

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

This edit checks for "usual" combinations of site and ICD-O-3 histology.

1. The Site/Histology validation list (available on the SEER web site, http://seer.cancer.gov/icd-o-3/) contains those histologies commonly found in the specified primary site. Histologies that occur only rarely or never are not included. These edits require review of all combinations not listed.
2. Since basal and squamous cell carcinomas of non-genital skin sites are not reportable to NPCR, these site/histology combinations do not appear on the SEER validation list.

Review of these cases requires investigating whether a) the combination is biologically implausible, or b) there are cancer registry coding conventions that would dictate different codes for the diagnosis. Review of these rare combinations often results in changes to the primary site and/or morphology, rather than a decision that the combination is correct.

Instructions for Coding

1. The data item is left blank if the program does not generate an error message for the edit Primary Site, Morphology-Type, Behavior ICDO3 (SEER IF25).
2. Any identified errors should have been corrected if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.

3. Code 1 indicates that the case was reviewed and both the site and histology are correct.

Codes

1	Reviewed and confirmed as reported
Blank	Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a primary site, histology, and behavior combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Histology

Over-ride Histology/Behavior Inter-field Review

SAS Alternate Name	Length	Source of Standard
I2040_ORHist	1	NAACCR Item #2040; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Diagnostic Confirmation, Behavior ICDO3 (SEER IF31)	Identifies records with a behavior of in situ and a non-microscopic diagnostic confirmation
Morphology--Type/Behavior ICDO3 (SEER MORPH)	

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flags as Used in the EDITS Software Package

The edit Diagnostic Confirmation, Behavior checks that, for *in situ* cases (Behavior = 2), Diagnostic Confirmation specifies microscopic confirmation (1, 2, or 4).

The distinction between *in situ* and invasive is very important to a registry, since prognosis is so different. Since the determination that a neoplasm has not invaded surrounding tissues, i.e., *in situ*, is made microscopically, cases coded *in situ* in behavior should have a microscopic confirmation code. However, very rarely, a physician will designate a case noninvasive or *in situ* without microscopic evidence.

The edit Morphology--Type/Behavior performs the following check:

1. Codes listed in ICD-O-3 with behavior codes of only 0 or 1 are considered valid, since the behavior matrix allows for the elevation of the behavior of such histologies when the tumor is *in situ* or malignant. This edit forces review of these rare cases to verify that they are indeed *in situ* or malignant.
2. The following ICD-O-3 histologies are generally not accepted as *in situ*: 8000-8005, 8020, 8021, 8331, 8332, 8800-9055, 9062, 9082, 9083, 9110-9493, 9501-9989. This edit forces review of these cases.

3. If a Morphology-Type/Behavior edit produces an error or warning message and the case is one in which the 4-digit morphology code is one that appears in ICD-O-3 only with behavior codes of 0 or 1, or the case is one in which the 4-digit morphology code is not generally accepted with a behavior code of 2, this edit forces review to verify the coding of morphology and that the behavior should be coded malignant or *in situ*.

   Exceptions:
   If year of Date of Diagnosis > 2000, then a behavior code of 1 is valid for the following ICD-O-3 histologies are valid with a behavior code of 1: 8442, 8451, 8462, 8472, and 8473.

   If year of Date of Diagnosis > 2003, the following ICD-O-3 benign histologies will pass without review: 8146, 8271, 8861, 8897, 9121, 9122, 9131, 9161, 9350, 9351, 9352, 9360, 9361, 9383, 9384, 9394, 9412, 9413, 9444, 9492, 9493, 9506, 9531, 9532, 9533, 9534, 9537, 9541, 9550, 9562, and 9570.
4. Grades 5-8 with histologies not in the range of 9590-9948 are impossible.
5. Some terms in ICD-O-3 carry an implied statement of grade. These histologies must be reported with the correct grade as stated below. An error of this type cannot be over-ridden.

   ICD-O-3
   8020/34 Carcinoma, undifferentiated
   8021/34 Carcinoma, anaplastic
   8331/31 Follicular adenocarcinoma, well differentiated
   9082/34 Malignant teratoma, undifferentiated
   9083/32 Malignant teratoma, intermediate type
   9401/34 Astrocytoma, anaplastic
   9451/34 Oligodendroglioma, anaplastic
   9511/31 Retinoblastoma, differentiated
   9512/34 Retinoblastoma, undifferentiated

Instructions for Coding

1. Tee data item is left blank if the program does not generate an error message for the edit Diagnostic Confirmation, Behavior ICDO3 (SEER IF31) or Morphology--Type/Behavior ICDO3 (SEER MORPH).
2. Any identified errors should have been corrected if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
3. Code 1, 2, or 3 indicates that the case was reviewed and confirms that the data are correct.

Codes

1 Reviewed and confirmed that the pathologist states the primary to be "*in situ*" or "malignant" although the behavior code of the histology is designated as "benign" or "uncertain" in ICD-O-2 or ICD-O-3

2 Reviewed and confirmed that the behavior code is "*in situ*," but the case is not microscopically confirmed
3 Reviewed and confirmed that conditions 1 and 2 both apply
Blank Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a diagnostic confirmation, histology, and behavior combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Report Source

Over-ride Type of Reporting Source/Sequence Number Inter-field Review (Inter-field Edit 04)

SAS Alternate Name	Length	Source of Standard
I2050_ORRptSrc	1	NAACCR Item #2050; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Type of Rep Srce(DC),Seq Num--Cent, ICDO3 (SEER IF04)	Identifies records with multiple primary cancers where one is reported only through a death certificate and histology code is <9590

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

The edit Type of Rep Srce(DC), Seq Num—Cent checks that if the case is a death-certificate-only case and the histology is not a lymphoma, leukemia, immunoproliferative, or myeloproliferative disease (ICD-O-3 histology is less than 9590), then the tumor sequence number must specify one primary only (sequence ‘00’).

Instructions for Coding

1. The data item is left blank if the program does not generate an error message for the reporting source edit.
2. Any identified errors should have been corrected if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
3. Code 1 indicates that the case was reviewed, confirms that the type of reporting source, histologic type, and tumor sequence number data are correct, verifies that a second or subsequent primary with a reporting source of death-certificate-only has been reviewed and is indeed an independent primary.

Codes

1 Reviewed and confirmed as reported

Blank Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a type of reporting source, histologic type, and tumor sequence number combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Ill-define Site

Over-ride Sequence Number/Ill-defined Site Inter-field Review (Inter-field Edit 22)

SAS Alternate Name	Length	Source of Standard
I2060_ORIlldefineSite	1	NAACCR Item #2060; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Seq Num--Central, Prim Site, Morph ICDO3 (SEER IF22)	Identifies records with multiple primary cancers where one is reported as an ill-defined primary site

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

The edit forces review of multiple primary cancers when one of the primaries is coded to a site/morphology combination that could indicate a metastatic site rather than a primary site.

1. If Sequence Number-Central indicates the person has had more than one primary, then any case with one of the following site/histology combinations requires review:
    - C760-C768 (ill-defined sites) or C809 (unknown primary) and ICD-O-3 histology < 9590.
    - C770-C779 (lymph nodes) and ICD-O-3 histology not in the range 9590-9729; or C420-C424 and ICD-O-3 histology not in the range 9590-9989. That combination is most likely a metastatic lesion.
    - Any site ICD-O-3 histology in the range 9740-9758.
2. If it turns out that the suspect tumor is a manifestation of one of the patient's other cancers, the metastatic or secondary case is deleted, remaining cases are re-sequenced, and the coding on the original case is corrected as necessary.

Instructions for Coding

1. The data item is left blank if the program does not generate an error message for the ill-defined primary site edit.

2. Any identified errors should have been corrected if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
3. Code 1 indicates that the case was reviewed and confirms that a second or subsequent primary report with an ill-defined primary site is indeed an independent primary.

Codes

1	Reviewed and confirmed as reported: a second or subsequent primary reported with an ill-defined primary site (C76.0-C76.8, C80.9) has been reviewed and is an independent primary
Blank	Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a primary site, histologic type, and tumor sequence number combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Leuk, Lymphoma

Over-ride Leukemia or Lymphoma/Diagnostic Confirmation Inter-field Review (Inter-field Edit 48)

SAS Alternate Name	Length	Source of Standard
I2070_ORLeukLymph	1	NAACCR Item #2070; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Diagnostic Confirmation, Histology ICDO3 (SEER IF48)	Identifies leukemia and lymphoma records where the diagnostic confirmation is not microscopic

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

Since lymphoma and leukemia are almost exclusively microscopic diagnoses, this edit forces review of any cases of lymphoma records that have a diagnostic confirmation of direct visualization or clinical, and any leukemia with a diagnostic confirmation of direct visualization.

Instructions for Coding

1. Thee data item is left blank if the program does not generate an error message for the Diagnostic Confirmation, Histology edit.
2. Any identified errors should have been corrected if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
3. Code 1 indicates that the case was reviewed and confirms that the histologic type and diagnostic confirmation are correctly coded. Positive hematologic findings and bone marrow specimens are included as histologic confirmation (code 1 in Diagnostic Confirmation) for leukemia**.**

Codes

1 Reviewed and confirmed as reported
Blank Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a diagnostic confirmation and histologic type combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Site/Behavior

Over-ride Flag for Site/Behavior (IF39)

SAS Alternate Name	Length	Source of Standard
I2071_ORSiteBehav	1	NAACCR Item #2071; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Primary Site, Behavior Code ICDO3 (SEER IF39)	Identifies records with a non-specific primary cancer site code with an *in situ* behavior

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See NAACCR Standards for Cancer Registries Volume II: Data Standards and Data Dictionary Chapter IV Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

The edit, Primary Site, Behavior Code, requires review of the following primary sites with a behavior of *in situ* (ICD-O-2 or ICD-O-3 behavior = 2):

C269 Gastrointestinal tract, NOS
C399 Ill-defined sites within respiratory system
C559 Uterus, NOS
C579 Female genital tract, NOS
C639 Male genital organs, NOS
C689 Urinary system, NOS
C729 Nervous system, NOS
C759 Endocrine gland, NOS
C760-C768 Ill-defined sites
C809 Unknown primary site

Since the designation of *in situ* is very specific and almost always requires microscopic confirmation, ordinarily specific information should also be available regarding the primary site. Conversely, if inadequate information is available to determine a specific primary site, it is unlikely that information about a cancer being *in situ* is reliable.

If an *in situ* diagnosis is stated, more specific primary site information should be sought. A primary site within an organ system can sometimes be identified based on the diagnostic

procedure or treatment given or on the histologic type. When no more specific site can be determined, a behavior code of 3 is usually assigned. In the exceedingly rare situation in which it is certain that the behavior is *in situ* and no more specific site code is applicable, Over-ride Site/Behavior is set to 1.

Instructions for Coding

1. The data item is left blank if the program does not generate an error message for the Primary Site, Behavior Code ICDO3 (SEER IF39) edit.
2. Any identified errors should have been corrected if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
3. Code 1 indicates that the case was reviewed and confirms that the *in situ* behavior and nonspecific site are correctly coded and that no further information about the primary site is available.

Codes

1	Reviewed and confirmed as reported
Blank	Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a primary site and behavior combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Over-ride Flags Data Items
Over-ride Site/Lat/Morph

Over-ride for Site/Laterality/Morphology (IF42)

SAS Alternate Name	Length	Source of Standard
I2074_ORSiteLatMorph	1	NAACCR Item #2074; SEER

Description

Some computer edits identify errors. Others indicate possible errors that require manual review for resolution. To eliminate the need to review the same cases repeatedly, over-ride flags have been developed to indicate that data in a record (or records) have been reviewed and, while unusual, are correct.

This over-ride is used with the following edits in the NAACCR Metafile of the EDITS software:

Inter-field Edit	Description
Laterality, Primary Site, Morph ICDO3 (SEER IF42)	Identifies records with a paired organ as the primary cancer site code with an in situ behavior and laterality is not coded to 1, 2, or 3.

Rationale

Some edits check for code combinations that are possible, but quite rare. If the code combination generates an error message and review of the case indicates that the codes are correct for the case, then the over-ride flag is used to skip the edit in the future. See Chapter IV, Recommended Data Edits and Software Coordination of Standards.

Over-ride Flag as Used in the EDITS Software Package

The edit Laterality, Primary Site, Morph requires that if the Primary Site is a paired organ and ICD-O-3 behavior is *in situ* (2), then laterality must be 1, 2, or 3.

The intent of this edit is to force review of *in situ* cases for which laterality is coded 4 (bilateral) or 9 (unknown laterality) as to origin. In rare instances when the tumor is truly midline (9) or the rare combination is otherwise confirmed correct, code 1 is entered for Override Site/Lat/Morph.

Instructions for Coding

- The data item is left blank if the program does not generate an error message for the Laterality, Primary site, Morph ICDO3 (SEER IF42) edit.
- Any identified errors should have been corrected if an item is discovered to be incorrect. The data item should be blank for records where identified errors were corrected.
- Code 1 indicates that the case was reviewed and confirms that the *in situ* behavior and laterality are correctly coded.

Codes

1 Reviewed and confirmed as reported
Blank Not reviewed or reviewed and corrected

Consideration for Use

Previous evaluations have shown that over-ride flags may be coded when not needed. These may result when edits are applied by individual state cancer registries that are more stringent than required for the CDC and SEER data submissions. Other instances may result from coding all over-ride flags for an individual record. If a data query results in a primary site and behavior combination that is unexpected, it is recommended that this data item be reviewed in conjunction with those records.

Section: Risk Factor-Associated Cancers Data Items
Alcohol-Related Cancers

Alcohol-Related Cancers

SAS Alternate Name	Source of Standard
Alcohol_related_cancers	Derived from NAACCR Items #400, #522, and #220

Description

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

Codes

See Appendix II for coding

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.
- 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*. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, see the Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors documentation, which includes references for each risk factor associated cancer category.
- 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).

Section: Risk Factor-Associated Cancers Data Items
HPV-Related Cancers

HPV-Related Cancers

SAS Alternate Name	Source of Standard
HPV_related_cancers	Derived from NAACCR Items #400, #522, #220 and #490

Description

Predefined variable created using ICD-O-3 site, histology and sex to define human papillomaviraus (HPV)-related cancers.

Codes

See Appendix II for coding

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.
- 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*. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, see the Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors documentation, which includes references for each risk factor associated cancer category.
- 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).

Section: Risk Factor-Associated Cancers Data Items
Obesity-Related Cancers

Obesity-Related Cancers

SAS Alternate Name	Source of Standard
Obesity_related_cancers	Derived from NAACCR Items #400, #522, #220, #490 and #230

Description

Predefined variable created using ICD-O-3 site, histology and sex to define obesity-related cancers.

Codes

See Appendix II for coding

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.
- 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*. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, see the Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors documentation, which includes references for each risk factor associated cancer category.
- 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).

Section: Risk Factor-Associated Cancers Data Items
Physical Inactivity-Related Cancers

Physical Inactivity-Related Cancers

SAS Alternate Name	Source of Standard
Physicalactivityrelated_cancers	Derived from NAACCR Items #400, #522, and #220

Description

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

Codes

See Appendix II for coding

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.
- 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*. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, see the Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors documentation, which includes references for each risk factor associated cancer category.
- 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).

Section: Risk Factor-Associated Cancers Data Items
Tobacco-Related Cancers

Tobacco-Related Cancers

SAS Alternate Name	Source of Standard
Tobacco_related_cancers	Derived from NAACCR Items #400, #522, and #220

Description

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

Codes

See Appendix II for coding

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.
- 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*. Then the number of *attributable* cancers can be estimated by multiplying the attributable fraction by the number of *associated* cancers.
- For more information, see the Predefined SEER*Stat Variables for Calculating the Number of Associated Cancers for Selected Risk Factors documentation, which includes references for each risk factor associated cancer category.
- 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).

Appendix I

(Only for example)

```
*** define race recode (w, b, ai/an, api) ***;
if I160_Race1 = '01' then racerec='1';        *** white;
else if I160_Race1 = '02' then racerec='2';   *** black;
else if I160_Race1 in ('03') then racerec = '3'; *** AI/AN;
else if I160_Race1 in ('98') then racerec = '5'; *** other unspecified;
else if I160_Race1 = '99' then racerec = '9';         *** unknown;
else if I160_Race1 in
('04','05','06','07','08','09','10','11','12','13','14','15','16','17','20','21','22','25','26','27','28',
                '30','31','32','96','97') then racerec = '4';          *** API;
else racerec=' ';

*** if white, check race2 ***;
if racerec='1' then do;
   if I161_Race2 = '02' then racerec='2';   *** black;
   else if I161_Race2 in ('03') then racerec = '3'; *** AI/AN;
   else if I161_Race2 in
('04','05','06','07','08','09','10','11','12','13','14','15','16','17','20','21','22','25','26','27','28',
                  '30','31','32','96','97') then  racerec = '4';       *** API;
end;

** AGE recode**;
/*** For I230_AgeDx, no recode group for value 999 ***/
if I230_AgeDx = '000' then AGEREC = '00';
else if I230_AgeDx >= '001' and I230_AgeDx <= '004' then AGEREC = '01';
else if I230_AgeDx >= '005' and I230_AgeDx <= '009' then AGEREC = '02';
else if I230_AgeDx >= '010' and I230_AgeDx <= '014' then AGEREC = '03';
else if I230_AgeDx >= '015' and I230_AgeDx <= '019' then AGEREC = '04';
else if I230_AgeDx >= '020' and I230_AgeDx <= '024' then AGEREC = '05';
else if I230_AgeDx >= '025' and I230_AgeDx <= '029' then AGEREC = '06';
else if I230_AgeDx >= '030' and I230_AgeDx <= '034' then AGEREC = '07';
else if I230_AgeDx >= '035' and I230_AgeDx <= '039' then AGEREC = '08';
else if I230_AgeDx >= '040' and I230_AgeDx <= '044' then AGEREC = '09';
else if I230_AgeDx >= '045' and I230_AgeDx <= '049' then AGEREC = '10';
else if I230_AgeDx >= '050' and I230_AgeDx <= '054' then AGEREC = '11';
else if I230_AgeDx >= '055' and I230_AgeDx <= '059' then AGEREC = '12';
else if I230_AgeDx >= '060' and I230_AgeDx <= '064' then AGEREC = '13';
else if I230_AgeDx >= '065' and I230_AgeDx <= '069' then AGEREC = '14';
else if I230_AgeDx >= '070' and I230_AgeDx <= '074' then AGEREC = '15';
else if I230_AgeDx >= '075' and I230_AgeDx <= '079' then AGEREC = '16';
else if I230_AgeDx >= '080' and I230_AgeDx <= '084' then AGEREC = '17';
else if I230_AgeDx >= '085' and I230_AgeDx <= '120' then AGEREC = '18';
```

Appendix II

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
AGEREC	Age Recode	NPCR Age recode with <1 year old	Char	2	00=00 years 01=01-04 years 02=05-09 years 03=10-14 years 04=15-19 years 05=20-24 years 06=25-29 years 07=30-34 years 08=35-39 years 09=40-44 years 10=45-49 years 11=50-54 years 12=55-59 years 13=60-64 years 14=65-69 years 15=70-74 years 16=75-79 years 17=80-84 years 18=85+ years
ALCOHOL_RELATED_CANCERS	Alcohol-Related Cancers	Alcohol-related cancers	Num	8	0=Lip, oral cavity, & pharynx 1=Esophagus 2=Colon & rectum 3=Liver 4=Larynx 5=Female breast
AYA_SITE_RECODE_2020_REVISION	AYA Site Recode 2020 Revision	AYA site recode 2020 revision	Num	8	1=1.1 Acute lymphoblastic leukemia 2=1.2.1 Acute promyelocytic leukemia 3=1.2.2 Other acute myeloid leukemia 4=1.3 Chronic myeloid leukemia 5=1.4 Chronic lymphocytic leukemia 6=1.5 Polycythemia vera 7=1.6 Essential thrombocythemia 8=1.7 Primary myelofibrosis 9=1.8 Myelodysplastic syndrome (MDS) 10=1.9.1 Hairy cell leukemia

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					11=1.9.2 Other lymphocytic/lymphoblastic leukemias 12=1.9.3 Other myeloid leukemias 13=1.9.4 Leukemias of mixed phenotype 14=1.9.5 Mast cell diseases 15=1.9.6 Other 16=2.1.1 Lymphoblastic 17=2.1.2 Burkitt 18=2.1.3 Diffuse large B-cell (DLBCL) 19=2.1.4 Primary mediastinal large B-cell excluded from DLBCL 20=2.1.5 Anaplastic T- and null-cell excluding NK/T-cell 21=2.1.6 Follicular 22=2.1.7 NK/T-cell (excluded from anaplastic T-cell) 23=2.1.8 MALT (mucosa-associated lymphoid tissue) 24=2.1.9 Other non-Hodgkin lymphoma NOS 25=2.2.1 Hodgkin NLP 26=2.2.2 Hodgkin classic - other 27=2.3 Myeloma 28=2.4 Cutaneous lymphomas 29=2.5 Other B- and T-cell lymphomas 30=2.6.1 Histiocytic and dendritic cell neoplasms 31=2.6.2 Lymphoma NOS 32=3.1.1.1 Oligodendrioglioma - benign/borderline 33=3.1.1.2 Oligodendrioglioma - invasive 34=3.1.2.1 Gliofibroma - benign/borderline 35=3.1.2.2 Glioblastoma - invasive 36=3.1.3.1 Ependymoma - benign/borderline 37=3.1.3.2 Ependymoma - invasive 38=3.1.4.1 Pilocytic astrocytoma 39=3.1.4.2 Other astrocytoma/astroglial - benign/borderline 40=3.1.4.3 Other astrocytoma/astroglial - invasive 41=3.2 Medulloblastoma and other invasive embryonal CNS tumors 42=3.3.1 Ganglioneuroma - benign/borderline 43=3.3.2 Neuroblastoma/ganglioneuroblastoma - invasive 44=3.4.1 Neuronal and mixed neuronal-glial - benign/borderline

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					45=3.4.2 Neuronal and mixed neuronal-glial - invasive 46=3.5.1 Meningioma - benign/borderline 47=3.5.2 Meningioma - invasive 48=3.6.1 Choroid plexus - benign/borderline 49=3.6.2 Choroid plexus - invasive 50=3.7.1 Craniopharyngioma - benign/borderline 51=3.7.2 Craniopharyngioma - invasive 52=3.8.1 Pituitary - benign/borderline 53=3.8.2 Pituitary - invasive 54=3.9.1 Pineal - benign/borderline 55=3.9.2 Pineal - invasive 56=3.10.1 Other and unspecified CNS - benign/borderline 57=3.10.2 Other and unspecified CNS - invasive 58=4.1 Osteosarcoma 59=4.2 Chondrosarcoma 60=4.3.1 Bone 61=4.3.2 Soft tissue 62=4.4.1 Myxofibrosarcoma 63=4.4.2 Malignant fibrous histiocytoma 64=4.4.3 Other fibromatous neoplasms 65=4.5 Liposarcoma 66=4.6 Synovial sarcoma 67=4.7 Leiomyosarcoma 68=4.8 Rhabdomyosarcoma 69=4.9 Gastrointestinal stromal tumor, malignant 70=4.10 Spindle cell sarcoma 71=4.11 Epithelioid sarcoma 72=4.12 Desmoplastic small round cell tumor 73=4.13 Chordoma 74=4.14 Giant cell sarcoma 75=4.15 Other soft tissue sarcomas 76=4.16 Other bone tumors 77=5.1.1 Hemangioblastoma and tufted hemangioma 78=5.1.2 Cavernous hemangioma 79=5.1.3 Other 80=5.2.1 Kaposi sarcoma 81=5.2.2 Other 82=6.1.1 Neurilemmoma 83=6.1.2 Other

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					84=6.2.1.1 CNS 85=6.2.1.2 Peripheral 86=6.2.2 Other 87=7.1.1.1 Seminoma 88=7.1.1.2 Embryonal carcinoma 89=7.1.1.3 Endodermal sinus (yolk sac tumor) 90=7.1.1.4 Teratoma 91=7.1.1.5 Mixed germ cell 92=7.1.1.6 Choriocarcinoma and other trophoblastic 93=7.1.1.7 Other 94=7.1.2.1 Carcinoma 95=7.1.2.2 Sex cord 96=7.2.1.1 Teratoma 97=7.2.1.2 Dysgerminoma 98=7.2.1.3 Yolk sac 99=7.2.1.4 Mixed germ cell 100=7.2.1.5 Other germ cell and trophoblastic 101=7.2.2.1.1.1 Clear cell adenocarcinoma 102=7.2.2.1.1.2 Cystadenocarcinoma 103=7.2.2.1.1.3 Mixed cell adenocarcinoma 104=7.2.2.1.1.4 Mucinous adenocarcinoma 105=7.2.2.1.1.5 Endometrioid 106=7.2.2.1.1.6 Other adenocarcinoma 107=7.2.2.1.2 Other carcinoma 108=7.2.2.2 Sex cord and other specialized gonadal 109=7.3 Germ cell and trophoblastic - CNS 110=7.4 Germ cell and trophoblastic excluding CNS, ovary, testis 111=7.5 Non-germ cell specified tumors excluding CNS, ovary, testis 112=7.6 Fibroepithelial including Brenner excluding breast phyllodes 113=8.1 Superficial spreading/low cumulative sun damage melanoma 114=8.2 Nodular melanoma 115=8.3 Other malignant 116=9.1.1 Medullary 117=9.1.2 Hurthle cell carcinoma 118=9.1.3 Papillary

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					119=9.1.4 Follicular 120=9.1.5 Papillary with follicular variant 121=9.1.6 Other 122=9.2.1.1 Nasopharyngeal carcinoma - squamous 123=9.2.1.2 Nasopharyngeal carcinoma - other 124=9.2.2.1 Oral cavity, lip, and pharynx - squamous 125=9.2.2.2 Oral cavity, lip, and pharynx - mucoepidermoid 126=9.2.2.3 Oral cavity, lip, and pharynx - other 127=9.2.3.1 Salivary gland - acinar 128=9.2.3.2 Salivary gland - other malignant 129=9.2.4 Other carcinoma of head and neck 130=9.3.1 Carcinoma of esophagus 131=9.3.2.1.1 Neuroendocrine tumor (NET) 132=9.3.2.1.2 Neuroendocrine carcinoma (NEC) 133=9.3.2.2 Stomach - signet ring 134=9.3.2.3 Stomach - other adenocarcinoma 135=9.3.2.4 Stomach - other invasive 136=9.3.3.1.1 NET 137=9.3.3.1.2 NEC 138=9.3.3.2 Small intestine - other 139=9.3.4.1.1 NET 140=9.3.4.1.2 NEC 141=9.3.4.1.3 Appendix - other 142=9.3.4.2.1.1 NET 143=9.3.4.2.1.2 NEC 144=9.3.4.2.2 Colon excluding appendix - adenocarcinoma 145=9.3.4.2.3 Colon excluding appendix - other 146=9.3.5.1.1 NET 147=9.3.5.1.2 NEC 148=9.3.5.2 Rectum - adenocarcinoma 149=9.3.5.3 Rectum - other 150=9.3.6.1 Anus - squamous 151=9.3.6.2 Anus - other 152=9.3.7.1 Liver and IBD - cholangiocarcinoma 153=9.3.7.2 Liver and IBD - hepatocellular carcinoma 154=9.3.7.3 Liver and IBD - other

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					155=9.3.8 Carcinoma of gallbladder and other extrahepatic biliary 156=9.3.9.1.1 NET 157=9.3.9.1.2 NEC 158=9.3.9.1.3 Neuroendocrine - other 159=9.3.9.2 Pancreas - adenocarcinoma 160=9.3.9.3 Pancreas - other 161=9.3.10 Other carcinoma of gastrointestinal tract 162=9.4.1 Small cell carcinoma - neuroendocrine carcinoma (NEC) 163=9.4.2.1 Non-small cell - adenocarcinoma 164=9.4.2.2.1 Non-small cell NET 165=9.4.2.2.2 Non-small cell NEC 166=9.4.2.3 Non-small cell - other 167=9.5 Carcinoma of skin (if collected) 168=9.6.1 Breast - infiltrating duct 169=9.6.2 Breast - adenocarcinoma 170=9.6.3 Breast - lobular 171=9.6.4 Breast - phyllodes 172=9.6.5 Breast - medullary 173=9.6.6 Breast - Paget 174=9.6.7 Breast - ductal 175=9.6.8 Breast - metaplastic 176=9.6.9 Breast - inflammatory 177=9.6.10 Breast - other 178=9.7.1.1 Cervix - squamous 179=9.7.1.2 Cervix - adenosquamous 180=9.7.1.3 Cervix - adenocarcinoma 181=9.7.1.4 Cervix - other 182=9.7.2.1.1 Corpus uteri - endometrioid 183=9.7.2.1.2 Corpus uteri - other adenocarcinoma 184=9.7.2.2 Corpus uteri - other 185=9.7.3 Carcinoma of vulva and vagina 186=9.7.4 Carcinoma of penis 187=9.7.5 Carcinoma of prostate 188=9.7.6 Other genital 189=9.8.1.1.1 Kidney - renal cell 190=9.8.1.1.2 Kidney - other adenocarcinoma 191=9.8.1.2 Kidney - other

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					192=9.8.2.1 Urinary bladder - transitional cell carcinoma 193=9.8.2.2 Urinary bladder - other carcinoma 194=9.8.3 Other urinary 195=9.9.1 Adrenocortical carcinoma 196=9.9.2 Unknown primary 197=9.9.3 Thymic carcinoma 198=9.9.4 Carcinoma of other and ill-defined sites 199=10.1.1 Wilms tumor 200=10.1.2 Olfactory and other non-CNS neuroblastomas 201=10.1.3 Other neuronal and embryonal non-CNS tumors 202=10.2.1 Paraganglioma - non-CNS 203=10.2.2 Other specified neoplasms 204=11. Unspecified malignant neoplasms except CNS 205=A.1.1 Superficial spreading melanoma in situ 206=A.1.2 Lentigo maligna 207=A.1.3 Other in situ melanoma 208=A.2 Colon including appendix - in situ 209=A.3 Rectum - in situ 210=A.4 Anus - in situ 211=A.5 Breast - in situ 213=A.7 Ovary - in situ 214=A.8 Vulva and vagina - in situ 215=A.9 Penis - in situ 216=A.10 Prostate in situ including PIN III 217 = A.11 Urinary bladder - in situ 218=A.12 Other in situ 999=Unclassified
BEHAVIOR_RECODE_FOR_ANALYSIS	Behavior recode for analysis derived/WHO2008	Behavior recode for analysis derived/WHO 2008	Num	8	0=Benign 1=Borderline malignancy 2=In situ 3=Malignant 4=Only malignant in ICD-O-3 6=Only malignant 2010+ * Not available for Puerto Rico's cases diagnosed from July to December 2017
CENSUS_REGION	Address at Diagnosis – Census Region	Address at Diagnosis-Census Region	Char	9	Midwest Northeast

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					South West
DXYEAR	Year of diagnosis	Year of diagnosis	Char	4	1998-2020
ECON_STATUS	Economic Status	Econ status	Num	8	0=Distressed 1=At Risk 2=Transitional 3=Competitive 4=Attainment 9=Missing county data 14=Blank(s)
HPV_RELATED_CANCERS	HPV-Related Cancers	HPV-related cancers	Num	8	0=Oropharyngeal squamous cell carcinoma 1=Anal and rectal squamous cell carcinoma 2=Vulvar squamous cell carcinoma 3=Vaginal squamous cell carcinoma 4=Penile squamous cell carcinoma 5=Cervical carcinoma
I1290_RXSUMMSURGPRIMSITE	RX Summ—Surgery Primary Site	RX Summ - Surg prim site	Char	2	00=None 10-19=Site-specific code; tumor destruction 20-80=Site-specific codes; resection 90=Surgery, NOS 98=Site specific codes; special 99=Unknown Blank=Blank(s)
I160_RACE1	Race 1	Race 1	Char	2	01=White 02=Black 03=American Indian/Alaska Native 04=Chinese 05=Japanese 06=Filipino 07=Hawaiian 08=Korean (1988+) 10=Vietnamese (1988+) 11=Laotian (1988+) 12=Hmong (1988+) 13=Kampuchean (1988+) 14=Thai (1994+) 15=Asian Indian or Pakistani, NOS (1988+) 16=Asian Indian (2010+ code) 17=Pakistani (2010+ code)

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					20=Micronesian, NOS (1991+) 21=Chamorran (1991+) 22=Guamanian, NOS (1991+) 25=Polynesian, NOS (1991+) 26=Tahitian (1991+) 27=Samoan (1991+) 28=Tongan (1991+) 30=Melanesian, NOS (1991+) 31=Fiji Islander (1991+) 32=New Guinean (1991+) 96=Other Asian (1991+) 97=Pacific Islander, NOS (1991+) 98=Other 99=Unknown
I161_RACE2	Race 2	Race 2 [161]	Char	2	01=White 02=Black 03=American Indian, Aleutian, or Eskimo (includes all indigenous populations of the Western hemisphere) 04=Chinese 05=Japanese 06=Filipino 07=Hawaiian 08=Korean 10=Vietnamese 11=Laotian 12=Hmong 13=Kampuchean (Cambodian) 14=Thai 15=Asian Indian or Pakistani, NOS 16=Asian Indian 17=Pakistani 20=Micronesian, NOS 21=Chamorro/Chamoru 22=Guamanian, NOS 25=Polynesian, NOS 26=Tahitian 27=Samoan 28=Tongan 30=Melanesian, NOS

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					31=Fiji Islander 32=New Guinean 88=No further race documented 96=Other Asian, including Asian, NOS and Oriental, NOS 97=Pacific Islander, NOS 98=Other 99=Unknown Blank=Blank, Race 2 not coded
I192_IHS	IHS Link	IHS Link	Char	1	0=Record sent for linkage, no IHS match 1=Record sent for linkage, IHS match Blank=Blank(s)
I1990_ORAGESITEMORPH	Over-ride Age/Site/Histology Inter-field Review (Inter-field Edit 15)	Over-ride Age/Site/Morph [1990]	Char	1	1=Reviewed and confirmed that age/site/histology combination is correct as reported 2=Reviewed and confirmed that case was diagnosed in utero 3=Reviewed and confirmed that conditions 1 and 2 both apply Blank=Not reviewed or reviewed and corrected.
I2000_ORSEQNODXCONF	Over-ride Sequence Number/Diagnostic Confirmation Inter-field Review (Inter-field Edit 23)	Over-ride SeqNo/DxConf [2000]	Char	1	1=Reviewed and confirmed as reported Blank=Not reviewed or reviewed and corrected
I2010_ORSITELATSEQNO	Over-ride Site/Histology/Laterality/Sequence Number Inter-record Review (Inter-record Edit 09)	Over-ride Site/Lat/SeqNo [2010]	Char	1	1=Reviewed and confirmed as reported Blank=Not reviewed or reviewed and corrected
I2030_ORSITETYPE	Over-ride Site/Type Inter-field Review (Inter-field Edit 25)	Over-ride Site/Type [2030]	Char	1	1=Reviewed and confirmed as reported Blank=Not reviewed or reviewed and corrected
I2040_ORHIST	Over-ride Histology/Behavior Inter-field Review	Over-ride Histology [2040]	Char	1	1=Reviewed and confirmed that the pathologist states the primary to be "in situ" or "malignant" although the behavior code of the histology is designated as "benign" or "uncertain" in ICD-O-2 or ICD-O-3 2=Reviewed and confirmed that the behavior code is "in situ," but the case is not microscopically confirmed 3=Reviewed and confirmed that conditions 1 and 2 both apply Blank=Not reviewed or reviewed and corrected

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
I2050_ORRPTSRC	Over-ride Type of Reporting Source/Sequence Number Inter-field Review (Inter-field Edit 04)	Over-ride Report Source [2050]	Char	1	1=Reviewed and confirmed as reported Blank=Not reviewed or reviewed and corrected
I2060_ORILLDEFINESITE	Over-ride Sequence Number/Ill-defined Site Inter-field Review (Inter-field Edit 22)	Over-ride Ill-define Site [2060]	Char	1	1=Reviewed and confirmed as reported: a second or subsequent primary reported with an ill-defined primary site (C76.0-C76.8, C80.9) has been reviewed and is an independent primary Blank=Not reviewed or reviewed and corrected
I2070_ORLEUKLYMPH	Over-ride Leukemia or Lymphoma/Diagnostic Confirmation Inter-field Review (Inter-field Edit 48)	Over-ride Leuk, Lymphoma [2070]	Char	1	1=Reviewed and confirmed as reported Blank=Not reviewed or reviewed and corrected
I2071_ORSITEBEHAV	Over-ride Flag for Site/Behavior (IF39)	Over-ride Site/Behavior [2071]	Char	1	1=Reviewed and confirmed as reported Blank=Not reviewed or reviewed and corrected
I2074_ORSITELATMORPH	Over-ride for Site/Laterality/Morphology (IF42)	Over-ride Site/Lat/Morph [2074]	Char	1	1=Reviewed and confirmed as reported Blank=Not reviewed or reviewed and corrected
I220_SEX	Sex	Sex	Char	1	1=Male 2=Female
I230_AGEDX	Age at Diagnosis	Age at diagnosis	Char	3	000-099 99+ =Age>99 years 999 = unknown
I240_DOB	Birth Date	Year of birth of the patient	Char	4	4-digit character 9999=(Age=999 or >99 years)
I2880_CSSSF1	CS Site Specific Factor 1	CS site-specific factor 1	Char	3	010-999=010-999 Blank=Blank(s)
I3020_DERIVEDSS2000	Derived SS2000	Derived SS2000	Char	1	0=IS 1=L 2=RE 3=RN 4=RE+RN 5=RNOS 7=D 8=NA 9=U Blank=Blank(s)

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
I380_SEQNOCNTRL	Sequence Number – Central	Sequence number - central	Char	2	00=One primary only 01=1st of 2 or more primaries 02=2nd of 2 or more primaries 03=3rd of 3 or more primaries 04=4th of 4 or more primaries 05=5th of 5 or more primaries 06=6th of 6 or more primaries 07=7th of 7 or more primaries 08=8th of 8 or more primaries 09=9th of 9 or more primaries 10=10th of 10 or more primaries 11=11th of 11 or more primaries 12=12th of 12 or more primaries 13=13th of 13 or more primaries 14=14th of 14 or more primaries 15=15th of 15 or more primaries 16=16th of 16 or more primaries 17=17th of 17 or more primaries 18=18th of 18 or more primaries 19=19th of 19 or more primaries 20=20th of 20 or more primaries 21=21st of 21 or more primaries 22=22nd of 22 or more primaries 23=23rd of 23 or more primaries 24=24th of 24 or more primaries 25=25th of 25 or more primaries 26=26th of 26 or more primaries 27=27th of 27 or more primaries 28=28th of 28 or more primaries 29=29th of 29 or more primaries 30=30th of 30 or more primaries 31=31st of 31 or more primaries 32=32nd of 32 or more primaries 33=33rd of 33 or more primaries 34=34th of 34 or more primaries 35=35th of 35 or more primaries 36=36th of 36 or more primaries 37=37th of 37 or more primaries 38=38th of 38 or more primaries

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					39=39th of 39 or more primaries 40=40th of 40 or more primaries 41=41st of 41 or more primaries 42=42nd of 42 or more primaries 43=43rd of 43 or more primaries 44=44th of 44 or more primaries 45=45th of 45 or more primaries 46=46th of 46 or more primaries 47=47th of 47 or more primaries 48=48th of 48 or more primaries 49=49th of 49 or more primaries 50=50th of 50 or more primaries 51=51st of 51 or more primaries 52=52nd of 52 or more primaries 53=53rd of 53 or more primaries 54=54th of 54 or more primaries 55=55th of 55 or more primaries 56=56th of 56 or more primaries 57=57th of 57 or more primaries 59=59th of 59 or more primaries 60=Only one state registry-defined neoplasm 61=1st of 2 or more state registry-defined neoplasms 62=2nd of 2 or more state registry-defined neoplasms 63=3rd of 3 or more state registry-defined neoplasms 64=4th of 4 or more state registry-defined neoplasms 65=5th of 5 or more state registry-defined neoplasms 66=6th of 6 or more state registry-defined neoplasms 67=7th of 7 or more state registry-defined neoplasms 68=8th of 8 or more state registry-defined neoplasms 69=9th of 9 or more state registry-defined neoplasms 70=10th of 10 or more state registry-defined neoplasms 71=11th of 11 or more state registry-defined neoplasms 72=12th of 12 or more state registry-defined neoplasms 75=15th of 15 or more state registry-defined neoplasms 80=20th of 20 or more state registry-defined neoplasms 87=27th of 20 or more state registry-defined neoplasms 88=Unknown seq num - state registry-defined neoplasms 98=Carcinoma in situ of the Cervix diagnosed 1/1/1996 or later

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					99=Unknown seq num - federally required in situ or malig tumors
I3843_GRADECLINICAL	Grade Clinical	Grade Clinical	Char	1	1,2,3,4,5,8,9 A,B,C,D,E,L,H,M,S Blank=Blank(s)
I3844_GRADEPATHOLOGICAL	Grade Pathological	Grade Pathological	Char	1	1,2,3,4,5,8,9 A,B,C,D,E,L,H,M,S Blank=Blank(s)
I3845_GRADEPOSTTHERAPY	Grade Post Therapy Path	Grade Post Therapy Path	Char	1	1,2,3,4,5,8,9 A,B,C,D,E,L,H,M Blank=Blank(s)
I390_DATEDX	Date of Diagnosis	Year and month of initial diagnosis	Char	6	6-digit character YYYYMM, month may be blank (Blank(s) or Invalid Value(s))
I410_LATERALITY	Laterality	Laterality	Char	1	0=Not a paired site 1=Right - origin of primary 2=Left - origin of primary 3=Only one side - side unspecified 4=Bilateral, single primary 5=Paired site: midline tumor 9=Paired site, but no information concerning laterality
I440_GRADE	Grade	Grade	Char	1	1=Well differentiated; Grade I 2=Moderately differentiated; Grade II 3=Poorly differentiated; Grade III 4=Undifferentiated; anaplastic; Grade IV 5=T-cell 6=B-cell; pre-B; B-precursor 7=Null cell; non T-non B 8=NK cell; natural killer cell (1995+) 9=Unknown Blank=Blank(s)
I490_DXCONF	Diagnostic Confirmation	Diagnostic Confirmation	Char	1	1=Positive histology 2=Positive exfoliative cytology, no positive histology 3=Pos hist AND immunophenotyping AND/OR pos genetic studies 4=Positive microscopic confirm, method not specified 5=Positive laboratory test/marker study 6=Direct visualization without microscopic confirmation 7=Radiography without microscopic confirm 8=Clinical diagnosis only

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					9=Unknown
I500_TYPERPTSRC	Type of Reporting Source	Type of Reporting Source	Char	1	1=Hospital inpatient/outpatient or clinic 2=Radiation treatment or medical oncology center (2006+) 3=Laboratory only (hospital or private) 4=Physicians office/private medical practitioner (LMD) 5=Nursing/convalescent home/hospice 6=Autopsy only 7=Death certificate only 8=Other hospital outpatient unit or surgery center (2006+) Blank=Blank(s)
I522_HISTTYPEICDO3	Histologic Type ICD-O-3	Histologic Type ICD-O-3	Char	4	8000-9992
I523_BEHAVICDO3	Behavior Code ICD-O-3	Behavior Code ICD-O-3	Char	1	0=Benign 1=Borderline malignancy 2=In situ 3=Malignant
I759_SS2000	SEER Summary Stage 2000	SEER Summary Stage 2000	Char	1	0=In situ 1=Localized only 2=Regional by direct extension only 3=Regional lymph nodes involved only 4=Regional by both direct extension and lymph node involvement 5=Regional, NOS 7=Distant site(s)/node(s) involved 8=Not applicable 9=Unknown/unstaged/unspecified/DCO Blank=Blank(s)
I760_SS1977	SEER Summary Stage 1977	SEER Summary Stage 1977	Char	1	0=In situ 1=Localized only 2=Regional by direct extension only 3=Regional lymph nodes involved only 4=Regional by both direct extension and lymph node involvement 5=Regional, NOS 7=Distant site(s)/node(s) involved 8=Not applicable 9=Unknown/unstaged/unspecified/DCO Blank=Blank(s)

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
I764_SS2018	Summary Stage 2018	Summary Stage 2018	Char	1	0=In situ 1=Localized only 2=Regional by direct extension only 3=Regional lymph nodes only 4=Regional by BOTH direct extension AND lymph node involvement 7=Distant site(s)/node(s) involved 8=Benign/borderline 9=Unknown Blank=Blank(s)
I80_STATEDX	Address at Diagnosis – State	Addr at DX - State	Char	2	Two letter USPS abbreviations for the 50 states, District of Columbia (DC), and Puerto Rico (PR).
I89_COUNTYDXANALYSIS	Address at Diagnosis – County	County at DX Analysis	Char	3	001-997=Valid FIPS code 998=Known town, city, state, or country of residence but county code not known AND a resident outside of the state of reporting institution (must meet all criteria). Use for Canadian residents. 999=The county of the patient is unknown, or the patient is not a United States resident. County is not documented in the patient's medical record. County is suppressed for Kansas, Minnesota for all dxyears, suppressed for Virginia in 2017-2020.
ICCCSITERECEXTENDEDICDO3WHO2008	ICCC site recode extended ICD-O-3/WHO 2008	ICCC site rec extended ICD-O-3/WHO 2008	Num	8	1=I(a.1) Precursor cell leukemias 2=I(a.2) Mature B-cell leukemias 3=I(a.3) Mature T-cell and NK cell leukemias 4=I(a.4) Lymphoid leukemia, NOS 5=I(b) Acute myeloid leukemias 6=I(c) Chronic myeloproliferative diseases 7=I(d) Myelodysplastic syndrome and other myeloproliferative 8=I(e) Unspecified and other specified leukemias 9=II(a) Hodgkin lymphomas 10=II(b.1) Precursor cell lymphomas 11=II(b.2) Mature B-cell lymphomas except Burkitt lymphoma 12=II(b.3) Mature T-cell and NK-cell lymphomas 13=II(b.4) Non-Hodgkin lymphomas, NOS 14=II(c) Burkitt lymphoma 15=II(d) Miscellaneous lymphoreticular neoplasms

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					16=II(e) Unspecified lymphomas 17=III(a.1) Ependymomas 18=III(a.2) Choroid plexus tumor 19=III(b) Astrocytomas 20=III(c.1) Medulloblastomas 21=III(c.2) PNET 22=III(c.3) Medulloepithelioma 23=III(c.4) Atypical teratoid/rhabdoid tumor 24=III(d.1) Oligodendrogliomas 25=III(d.2) Mixed and unspecified gliomas 26=III(d.3) Neuroepithelial glial tumors of uncertain orig 27=III(e.1) Pituitary adenomas and carcinomas 28=III(e.2) Tumors of sellar region (craniopharyngiomas) 29=III(e.3) Pineal parenchymal tumors 30=III(e.4) Neuronal and mixed neuronal-glial tumors 31=III(e.5) Meningiomas 32=III(f) Unspecified intracranial and intraspinal neoplasms 33=IV(a) Neuroblastoma and ganglioneuroblastoma 34=IV(b) Other peripheral nervous cell tumors 35=V Retinoblastoma 36=VI(a.1) Nephroblastoma 37=VI(a.2) Rhabdoid renal tumor 38=VI(a.3) Kidney sarcomas 39=VI(a.4) pPNET of kidney 40=VI(b) Renal carcinomas 41=VI(c) Unspecified malignant renal tumors 42=VII(a) Hepatoblastoma 43=VII(b) Hepatic carcinomas 44=VII(c) Unspecified malignant hepatic tumors 45=VIII(a) Osteosarcomas 46=VIII(b) Chondrosarcomas 47=VIII(c.1) Ewing tumor and Askin tumor of bone 48=VIII(c.2) pPNET of bone 49=VIII(d.1) Malignant fibrous neoplasms of bone 50=VIII(d.2) Malignant chordomas 51=VIII(d.3) Odontogenic malignant tumors 52=VIII(d.4) Miscellaneous malignant bone tumors 53=VIII(e) Unspecified malignant bone tumors

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					54=IX(a) Rhabdomyosarcomas 55=IX(b.1) Fibroblastic and myofibroblastic tumors 56=IX(b.2) Nerve sheath tumors 57=IX(b.3) Other fibromatous neoplasms 58=IX(c) Kaposi sarcoma 59=IX(d.1) Ewing tumor and Askin tumor of soft tissue 60=IX(d.2) pPNET of soft tissue 61=IX(d.3) Extrarenal rhabdoid tumor 62=IX(d.4) Liposarcomas 63=IX(d.5) Fibrohistiocytic tumors 64=IX(d.6) Leiomyosarcomas 65=IX(d.7) Synovial sarcomas 66=IX(d.8) Blood vessel tumors 67=IX(d.9) Osseous & chondromatous neoplasms of soft tissue 68=IX(d.10) Alveolar soft parts sarcoma 69=IX(d.11) Miscellaneous soft tissue sarcomas 70=IX(e) Unspecified soft tissue sarcomas 71=X(a.1) Intracranial & intraspinal germinomas 72=X(a.2) Intracranial & intraspinal teratomas 73=X(a.3) Intracranial & intraspinal embryonal carcinomas 74=X(a.4) Intracranial & intraspinal yolk sac tumor 75=X(a.5) Intracranial & intraspinal choriocarcinoma 76=X(a.6) Intracranial & intraspinal tumors of mixed forms 77=X(b.1) Germinomas: extracranial/extragonadal 78=X(b.2) Malignant teratomas: extracranial/extragonadal 79=X(b.3) Embryonal carcinomas: extracranial/extragonadal 80=X(b.4) Yolk sac tumor: extracranial/extragonadal 81=X(b.5) Choriocarcinomas: extracranial/extragonadal 82=X(b.6) Other mixed germ cell: extracranial/extragonadal 83=X(c.1) Malignant gonadal germinomas 84=X(c.2) Malignant gonadal teratomas 85=X(c.3) Gonadal embryonal carcinomas 86=X(c.4) Gonadal yolk sac tumor 87=X(c.5) Gonadal choriocarcinoma

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					88=X(c.6) Malignant gonadal tumors of mixed forms 89=X(c.7) Malignant gonadal gonadoblastoma 90=X(d) Gonadal carcinomas 91=X(e) Other and unspecified malignant gonadal tumors 92=XI(a) Adrenocortical carcinomas 93=XI(b) Thyroid carcinomas 94=XI(c) Nasopharyngeal carcinomas 95=XI(d) Malignant melanomas 96=XI(e) Skin carcinomas 97=XI(f.1) Carcinomas of salivary glands 98=XI(f.2) Carcinomas of colon and rectum 99=XI(f.3) Carcinomas of appendix 100=XI(f.4) Carcinomas of lung 101=XI(f.5) Carcinomas of thymus 102=XI(f.6) Carcinomas of breast 103=XI(f.7) Carcinomas of cervix uteri 104=XI(f.8) Carcinomas of bladder 105=XI(f.9) Carcinomas of eye 106=XI(f.10) Carcinomas of other specified sites 107=XI(f.11) Carcinomas of unspecified site 108=XII(a.1) Gastrointestinal stromal tumor 109=XII(a.2) Pancreatoblastoma 110=XII(a.3) Pulmonary blastoma and pleuropulmonary blastoma 111=XII(a.4) Other complex mixed and stromal neoplasms 112=XII(a.5) Mesothelioma 113=XII(a.6) Other specified malignant tumors 114=XII(b) Other unspecified malignant tumors 253=Not classified by ICCC or in situ
LYMPHOID_NEOPLASM_RECODE_2021	Lymphoid neoplasm recode 2021 revision	Lymphoid neoplasm recode 2020	Num	8	1=1(a)1.1 Lymphocyte-rich 2=1(a)1.2 Mixed cellularity 3=1(a)1.3 Lymphocyte-depleted 4=1(a)2 Nodular sclerosis 5=1(a)3 Classical Hodgkin lymphoma, NOS 6=1(b) Nodular lymphocyte predominant Hodgkin lymphoma 7=2(a)1 Precursor Non-Hodgkin lymphoma, B-cell 8=2(a)2.1.1 Chronic/Small lymphocytic leuk/lymph

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					9=2(a)2.1.2 Prolymphocytic leukemia, B-cell 10=2(a)2.1.3 Mantle-cell lymphoma 11=2(a)2.2.1 Lymphoplasmacytic lymphoma 12=2(a)2.2.2 Waldenstrom macroglubulinemia 13=2(a)2.3.1 DLBCL, NOS 14=2(a)2.3.2 Intravascular large B-cell lymphoma 15=2(a)2.3.3 Primary effusion lymphoma 16=2(a)2.3.4 Mediastinal large B-cell lymphoma 17=2(a)2.4 Burkitt lymphoma/leukemia 18=2(a)2.5.1 Splenic MZL 19=2(a)2.5.2 Extranodal MZL, MALT type 20=2(a)2.5.3 Nodal MZL 21=2(a)2.6 Follicular lymphoma 22=2(a)2.7 Hairy-cell leukemia 23=2(a)2.8.1 Plasmacytoma 24=2(a)2.8.2 Multiple myeloma/plasma-cell leuk 25=2(a)2.9 Heavy chain disease 26=2(a)3 Non-Hodgkin lymphoma, B-cell, NOS 27=2(b)1 Precursor Non-Hodgkin lymphoma, T-cell 28=2(b)2.1.1 Mycosis fungoides 29=2(b)2.1.2 Sezary syndrome 30=2(b)2.2.1 Peripheral T-cell lymphoma, NOS 31=2(b)2.2.2 Angioimmunoblastic T-cell lymphoma 32=2(b)2.2.3 Subcutan panniculitis-like T-cell lymph 33=2(b)2.2.4 Anaplastic lar cell lymph, T-/Null-cell 34=2(b)2.2.5 Hepatosplenic T-cell lymphoma 35=2(b)2.2.6 Enteropathy-type T-cell lymphoma 36=2(b)2.2.7 Cutaneous T-cell lymphoma, NOS 37=2(b)2.2.8 Prim cutaneous anaplastic lar cell lymph 38=2(b)2.3 Adult T-cell leukemia/lymphoma 39=2(b)2.4 NK/T-cell lymph, nasal-type/aggres NK leuk 40=2(b)2.5 T-cell large granular lymphocytic leukemia 41=2(b)2.6 Prolymphocytic leukemia, T-cell 42=2(c) Non-Hodgkin lymphoma, unknown lineage 43=3 Composite Hodgkin lymphoma and NHL 44=4 Lymphoid neoplasm, NOS 61=Unclassified 63=Invalid value(s) PR

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
MERGED_ESTROGEN_RECEPTOR	Merged Estrogen Receptor	Merged estrogen receptor	Num	8	0=ER negative 1=ER positive 7=Test ordered, results not in chart 9=Not documented, indetermined, unknown 126=Blank(s)
MERGED_HER2_SUMMARY	Merged HER2 Summary	Merged HER2 summary	Num	8	0=HER2 negative, equivocal 1=HER2 positive 7=Test ordered, results not in chart 9=Not documented, indetermined, unknown 126=Blank(s)
MERGED_PROGESTERONE_RECEPTOR	Merged Progesterone Receptor	Merged progesterone receptor	Num	8	0=PR negative 1=PR positive 7=Test ordered, results not in chart 9=Not documented, indetermine, unknown 126=Blank(s)
MERGED_RADIATION	Merged Radiation	Merged Radiation	Num	8	1=had radiation 2=did not have radiation 9=unknown 14=Blank(s)
MERGED_SUMMARY_STAGE	Merged Summary Stage	Merged Summary Stage	Char	1	0=In situ 1=Localized only 2=Regional, direct extension only 3=Regional, regional lymph nodes only 4=Regional, direct extension and regional lymph nodes 5=Regional, NOS 6=Distant site(s)/node(s) involved 7=Benign/borderline 8=Not applicable 9=Unknown/unstaged/unspecified/DCO Blank=Blank(s) or invalid
NHIA_SUPP	State race ethnicity suppress		Char	1	0=non-Hispanic 1=Hispanic Blank=suppressed ethnicity
OBESITY_RELATED_CANCERS	Obesity-Related Cancers	Obesity-related cancers	Num	8	0=Esophageal adenocarcinoma 1=Gastric cardia 2=Colon & rectum 3=Liver 4=Gallbladder 5=Pancreas

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					6=Kidney 7=Meningioma 8=Thyroid 9=Multiple myeloma 10=Post-menopausal female breast cancer 11=Corpus & uterus NOS 12=Ovary
ORIGIN_RECODE_NHIA	Origin Recode NHIA (Hispanic, Non-Hisp)	Origin Recode NHIA (Hispanic, Non-Hisp)	Char	1	0=non-Hispanic 1=Hispanic 3=Unknown 9 = Suppressed - North Dakota and Wisconsin * Not available for Puerto Rico
PHYSICALACTIVITYRELATED_CANCERS	Physical Inactivity-Related Cancers	Physical activity-related cancers	Num	8	0=Colon 1=Post-menopausal female breast 2=Corpus and uterus NOS
PRIMARY_SITE	Primary Site	Primary Site (I400_Site)	Char	3	000-809
RACE_RECODE_W_B_AI_API	Race recode (W, B, AI, API)	Race recode (W, B, AI, API)	Num	8	1=White 2=Black 3=American Indian/Alaska Native 4=Asian or Pacific Islander 9=Unknown (including Other unspecified (1991+) * Not available for Puerto Rico
RACE_AND_ORIGIN_RECODE	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)	Num	8	1 = Non-Hispanic White 2 = Non-Hispanic Black 3 = Non-Hispanic American Indian/Alaska Native 4 = Non-Hispanic Asian or Pacific Islander 5 = Hispanic (All Races) 8 = Suppressed - North Dakota and Wisconsin 9 = Non-Hispanic Unknown Race * Not available for Puerto Rico
RACE_SUPP	State race ethnicity suppress		Char	1	1=no suppression by race for white 2=no suppression by race for black 3=no suppression by race of AI/AN (AI/AN suppressed states will be missing) 4=no suppression by race of API (API suppressed states will be missing) 5=no suppression by race for other Blank=suppressed race

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
RACEETH_SUPP	State race ethnicity suppress		Char	1	1=no suppression white NH 2=no suppression white HISP 3=no suppression black NH 4=no suppression black HISP 5=no suppression AIAN NH 6=no suppression AIAN HISP 7=no suppression API NH 8=no suppression API HISP Blank=suppressed race ethnicity
RANDOM_ID	Alternate Patient ID Number	Patient ID number	Char	8	8-digit character
RURALURBAN_CONTINUUM_2013	Rural-urban Continuum 2013	Ruralurban continuum 2013	Char	2	01=Counties in metropolitan areas ge 1 million pop 02=Counties in metropolitan areas of 250,000 to 1 million pop 03=Counties in metropolitan areas of lt 250 thousand pop 04=Urban pop of ge 20,000 adjacent to a metropolitan area 05=Urban pop of ge 20,000 not adjacent to a metropolitan area 06=Urban pop of 2,500 to 19,999, adjacent to a metro area 07=Urban pop of 2,500 to 19,999, not adjacent to a metro area 08=Comp rural lt 2,500 urban pop, adjacent to a metro area 09=Comp rural lt 2,500 urban pop, not adjacent to metro area 98=Valid state/county, but no corresponding rural/urban code 99=Missing or Unknown state/cnty info includes xx,yy,zz or 999
SITE_RECODE_ICD_O_3_WHO_2008	Site recode ICD-O-3/WHO 2008	Site recode ICD-O-3/WHO 2008	Num	8	1=Lip 2=Tongue 3=Salivary Gland 4=Floor of Mouth 5=Gum and Other Mouth 6=Nasopharynx 7=Tonsil 8=Oropharynx

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					9=Hypopharynx 10=Other Oral Cavity and Pharynx 11=Esophagus 12=Stomach 13=Small Intestine 15=Cecum 16=Appendix 17=Ascending Colon 18=Hepatic Flexure 19=Transverse Colon 20=Splenic Flexure 21=Descending Colon 22=Sigmoid Colon 23=Large Intestine, NOS 25=Rectosigmoid Junction 26=Rectum 27=Anus, Anal Canal and Anorectum 29=Liver 30=Intrahepatic Bile Duct 31=Gallbladder 32=Other Biliary 33=Pancreas 34=Retroperitoneum 35=Peritoneum, Omentum and Mesentery 36=Other Digestive Organs 37=Nose, Nasal Cavity and Middle Ear 38=Larynx 39=Lung and Bronchus 40=Pleura 41=Trachea, Mediastinum and Other Respiratory Organs 42=Bones and Joints 43=Soft Tissue including Heart 44=Melanoma of the Skin 45=Other Non-Epithelial Skin 46=Breast 47=Cervix Uteri 48=Corpus Uteri 49=Uterus, NOS 50=Ovary

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					51=Vagina 52=Vulva 53=Other Female Genital Organs 54=Prostate 55=Testis 56=Penis 57=Other Male Genital Organs 58=Urinary Bladder 59=Kidney and Renal Pelvis 60=Ureter 61=Other Urinary Organs 62=Eye and Orbit 63=Brain 64=Cranial Nerves Other Nervous System 65=Thyroid 66=Other Endocrine including Thymus 68=Hodgkin - Nodal 69=Hodgkin - Extranodal 71=NHL - Nodal 72=NHL - Extranodal 73=Myeloma 74=Acute Lymphocytic Leukemia 75=Chronic Lymphocytic Leukemia 76=Other Lymphocytic Leukemia 77=Acute Myeloid Leukemia 80=Acute Monocytic Leukemia 78=Chronic Myeloid Leukemia 89=Other Myeloid/Monocytic Leukemia 83=Other Acute Leukemia 85=Aleukemic, Subleukemic and NOS 87=Mesothelioma 88=Kaposi Sarcoma 86=Miscellaneous 127 = Invalid Value(s)
TOBACCO_RELATED_CANCERS	Tobacco-Related Cancers	Tobacco-related cancers	Num	8	0=Lip, oral cavity, & pharynx 1=Esophagus 2=Stomach 3=Colon & rectum 4=Liver

SAS VARIABLE NAME	VARIABLE NAME	LABEL	ATTR.	LENGTH	VALUES
					5=Pancreas 6=Larynx 7=Trachea, lung, & bronchus 8=Cervix uteri 9=Kidney & renal pelvis 10=Urinary bladder 11=Acute myeloid leukemia
USCS1120	USCS1120	Registry meet(1) USCS criteria for 2011-2020 years	Char	1	0=Does not meet USCS standard 1120 1=Meets USCS standard 1120
USCS1620	USCS1620	Registry meet(1) USCS criteria for 2016-2020 years	Char	1	0=Does not meet USCS standard 1620 1=Meets USCS standard 1620
USCS9820	USCS9820	Registry meet(1) USCS criteria for 1998-2020 years	Char	1	0=Does not meet USCS standard 9820 1=Meets USCS standard 9820
USCS9920	USCS9920	Registry meet(1) USCS criteria for 1999-2020 years	Char	1	0=Does not meet USCS standard 9920 1=Meets USCS standard 9920
USCSSTD	USCS Standard	Registry meet(1) USCS criteria for record specific DxYear	Char	1	0=Does not meet USCS standard 1=Meets USCS standard

Appendix III

State race ethnicity suppression

```
*******************************************
 define race recode (white, black, AI/AN, API)
*******************************************;
if I160_Race1 = '01' then racerec='1';        *** white;
else if I160_Race1 = '02' then racerec='2';   *** black;
else if I160_Race1 in ('03') then racerec = '3'; *** AI/AN;
else if I160_Race1 = ( '99',’98’) then racerec = '9';         *** unknown, other unspecified;;
else if I160_Race1 in ('04','05','06','07','08','09','10','11','12','13','14','15','16','17','20','21','22','25','26','27','28',
              '30','31','32','96','97') then  racerec = '4';          *** API;
else racerec=' ';


*******************************************
 if white, check race2
*******************************************;
if racerec='1' then do;
   if I161_Race2 = '02' then racerec='2';   *** black;
   else if I161_Race2 in ('03') then racerec = '3'; *** AI/AN;
   else if I161_Race2 in ('04','05','06','07','08','09','10','11','12','13','14','15','16','17','20','21','22','25','26','27','28',
                '30','31','32','96','97') then  racerec = '4';       *** API;
end;


*******************************************
 if white, check ihslink
*******************************************;

if racerec in ('1', '9') & I192_IHS='1' then racerec='3';

if I191_NHIA in ('0') then nhia='0'; *non-hispanic;
else if I191_NHIA in ('1','2','3','4','5','6','7','8') then nhia='1'; *hispanic;

/***create merged race/ethnicity;*/
if racerec='1' and nhia='0' then raceeth='1'; *white NH;
else if racerec='1' and nhia='1' then raceeth='2'; *white hisp;
else if racerec='2' and nhia='0' then raceeth='3'; *black NH;
else if racerec='2' and nhia='1' then raceeth='4'; *black hisp;
else if racerec='3' and nhia='0' then raceeth='5'; *aian NH;
else if racerec='3' and nhia='1' then raceeth='6'; *aian hisp;
else if racerec='4' and nhia='0' then raceeth='7'; *api NH;
else if racerec='4' and nhia='1' then raceeth='8'; *api hisp;

*******************************************
 create exclusion variables
*******************************************;

**** race exclusion ****;
  ** exclude only AI/AN from race **;
  if subm_st in ('KS','IL','NJ','NY') then exclude_aian=1;
  else exclude_aian=0;
```

```
  ** exclude only API from race **;
  if subm_st in ('KS') then exclude_api=1;
  else exclude_api=0;

 **** Race ethnicity exclusion *****;
  *** exclude from race-ethnicity (raceeth_supp) variable. We blank ethnicity (hisp, non-hisp for all races)***;
  if subm_st in ('ND','WI') then exclude_eth=1; *** all rases ***;


*****************************************************************
  Exclude states with race, ethnicity and NHIA (Hisp only supression)
  Custom supression (partial)
*****************************************************************;
  *** For partial race suppression for race-eth variable ***;
  else if (subm_st in ('IL','NJ','NY') and raceeth in ('5','6')) then exclude_eth=1; **** Hisp, non-hisp AI/AN
exclusion. Keep white, black, and API ***;

  else if (subm_st in ('KS') and raceeth in ('3','4','5','6','7','8')) then exclude_eth=1; **** selected races ethnicity
exclusion. Keep white only ***;
  else exclude_eth=0;



*************************************************************************************************
  create race and race/ethnicity variables to use for analysis - excludes states to be suppressed
*************************************************************************************************;

  ** Race recode **;
  if racerec in (1,2,5) then race_supp=racerec;                  ** no suppression by race for white, black, other **;
  else if racerec=3 & exclude_aian=0 then race_supp=racerec;      ** AI/AN suppressed states will be missing **;
  else if racerec=4 & exclude_api=0 then race_supp=racerec;       ** API suppressed states will be missing **;

  ** Race/ethnicity recode **;
  if exclude_eth=0 then raceeth_supp=raceeth;     ** suppressed states will be missing **;
  else raceeth_supp = " ";

  ** ethnicity alone **;
  if exclude_eth=0 then nhia_supp=nhia;           ** suppressed states will be missing **;
  else nhia_supp = " ";
```