U.S. Cancer Statistics
Restricted Access Data Set

Data Dictionary and Data Standards
2017 November Data Submission
Released June 2018

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

[Logos of CDC and NIH]
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.

Available Data

This file documents the data items included in the U.S. Cancer Statistics Restricted Access Data Set, 1998-2015.

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 2017 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, 2017), and
- SEER allowed an interval of 22 months after the close of the diagnosis year (data submission by November 1, 2017).

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 is as follows:
- 1998-2015: ~97.8%
- 1999-2015: ~97.8%
- 2001-2015: ~99.0%
- 2004-2015: 100%
- 2006-2015: 100%
- 2011-2015: 100%.

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2014</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2013</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2012</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2011</td>
<td>After manual review of Nevada’s data, all registries met the publication criteria, (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2010</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2009</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2008</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2007</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2006</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2005</td>
<td>All registries met the publication criteria (50 U.S. States, D.C., and Puerto Rico).</td>
</tr>
<tr>
<td>2004</td>
<td>All registries met the publication criteria (50 U.S. States and D.C.).</td>
</tr>
<tr>
<td>2003</td>
<td>All registries met the publication criteria (50 U.S. States and D.C.).</td>
</tr>
</tbody>
</table>
2002: All registries met the publication criteria except Mississippi; District of Columbia’s data was included after manual review. 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 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 Arkansas; data are not available for Mississippi and South Dakota. Counts and rates cover approximately 98% of the U.S. population (47 U.S. States and D.C.).

1999: All registries met the publication criteria except Arkansas; data are not available for Mississippi and South Dakota. Counts and rates cover approximately 98% of the U.S. population (47 U.S. States and D.C.).

1998: All registries met the publication criteria except Arkansas, Georgia, and New Hampshire; data are not available for Mississippi, South Dakota, and Tennessee. Counts and rates cover approximately 92% of the U.S. population (44 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. The adjusted US Census population estimates should be used; county-level populations in the four hurricane-affected states should be adjusted 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, analysts should read and understand the following section. If you have questions regarding these notes, please contact CDC at usesdata@cdc.gov.

State 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. States are also given the right to opt out of including their data in the dataset. Examine the table of state by year of diagnosis prior to beginning your analysis. See the Excel file referenced above. See the U.S. Cancer Statistics Eligibility Criteria section for more information on the criteria.

Four user-specified variables are included in the database: uscs9815, uscs9915, uscs0615, and uscs1115, 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, uscs1115 includes states that have data available for all five years, 2011 through 2015). 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

Complementary Cell Suppression

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

Suppressing less 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 less 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. More information can be found at:


**Reporting Delay**

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 outpatient 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 variable, `state race eth suppress`, 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 variable, `state race eth suppress`, is not used and state- or county-level data are being reported, the following suppressions must be made when using data from the 2017 submission:
Hispanic ethnicity data cannot be displayed for Arkansas, Delaware, Kentucky, and Massachusetts.

Data for American Indians and Alaska Natives (AI/AN) cannot be displayed for Delaware, Illinois, Kentucky, New Jersey, and New York.

Data for Asian and Pacific Islanders (API) cannot be displayed for Arkansas, Delaware, and Kentucky.

Any race and ethnicity combinations—for example, white Hispanic, white non-Hispanic, black Hispanic, black non-Hispanic—cannot be displayed for Arkansas, Delaware, Kentucky, Massachusetts and Pennsylvania.

Race Recode variable

This 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 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/ANs who are members of federally recognized tribes, estimated to be about 65% of the AI/AN population. To improve identification of AI/ANs, 32 NPCR registries with Purchase/Referred Care Service Delivery Area (PRCSDA) counties in their state annually link with the IHS patient registration database to identify AI/ANs who had not been coded as AI/AN (see “IHS Link” variable description). All NPCR registries link every five years. Linkages were performed by all NPCR registries most recently in 2011 and 2016 (cases diagnosed 1995-2014). In 2017, some, but not all, NPCR registries linked cancer cases diagnosed from 1995-2015; SEER registries linked cancer cases diagnosed from 1994-2015.

- 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 Contract health Services Delivery Areas (CHSDA) counties). States are given the right to suppress county-specific data every submission year.

Stage

Stage at diagnosis, or the extent to which the cancer has spread at the time it was originally diagnosed, is recorded using SEER Summary Stage 1977 through diagnosis year 2000, Summary Stage 2000 for diagnosis years 2001-2003, and the Collaborative Stage Data Collection System to derive Summary Stage 2000 for diagnosis years 2004-2015. As a general rule, Derived Summary Stage 2000 data should be analyzed for diagnosis years 2004-2015 in this file. Summary Stage 2000 should be used with 2001-2003 cases, and Summary Stage 1977 should be used with earlier cases.

In certain circumstances when information is not available to calculate the Derived Summary Stage 2000, NPCR permits the use of Summary Stage 2000 for diagnosis years 2004-2015. As a result, when stage information is required, it is recommended that the Merged Summary Stage 2000 variables described below be used to account for missing information in Derived Summary Stage 2000.

When the coding instructions between the three systems do not vary appreciably for a given primary site, a merged variable can be created and used for Summary Stage cases across 1998-2015. For primary sites where the coding instructions changed to redistribute the percentage of cases coded as localized, regional, and distant, analyses should be limited to 2001 cases and forward. See the NAACCR “Site-Specific Comparison of Summary Stage 1977 and Summary Stage 2000 Coding” for specific information.

A merged variable, Merged Summary Stage 2000, has been created to span two time periods when two different staging schemes were used. Stage at diagnosis is recorded using SEER Summary Stage 2000 for diagnosis years 2001-2003, Derived SEER Summary Stage 2000 for
diagnosis years 2004-2014, and either Summary Stage 2000 or Derived SEER Summary Stage 2000 for diagnosis year 2015.

The coding logic for this merged variable is:

- If a case was diagnosed between 2001 and 2003, stage at diagnosis is recorded using the SEER Summary Stage 2000 variable value.
- If a case was diagnosed between 2004 and later, 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 or unknown and a valid value is available for the SEER Summary Stage 2000 variable, that value is used to populate the merged variable. For example, if a case was diagnosed in 2013 and Derived SEER Summary Stage was blank or unknown, but SEER Summary Stage had a value of “local,” then the merged variable was coded as local stage. Otherwise, the merged variable is left blank or unknown for that record.

**Primary Site Variables**

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. These changes are incorporated into the site recode variables in this data set.

If the standard primary site recode variable is not used and a user-defined primary site variable is created, leukemias and lymphomas (9590-9989) should be excluded. 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/](http://seer.cancer.gov/siterecode/).

**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, your analysis should be restricted 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*. 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.

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]1% 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).


U.S. Cancer Statistics Eligibility Criteria

After years of analyzing completeness of case ascertainment, we have determined that our registries consistently deliver high quality, complete data. We have discontinued using

__________________________

1 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. For population counts, please see Appendix II.
completeness of case ascertainment calculations as a measure of eligibility for publication. The data quality criteria—missing/unknown data, death certificate only percentage, duplicate rate, and percent records passing edits—will continue to be used in determining inclusion. Even though the completeness estimate will no longer be a criterion for U.S. Cancer Statistics, we will continue to use this estimate to monitor and evaluate progress in meeting NPCR Program Standards.

Cancer registries in the “meets uscs criteria” category for each year and for groups of years at the Regional Level meet the following data quality criteria for all cancer sites combined:

- **No more than 5% of cases are ascertained solely on the basis of a death certificate.** The proportion of cases ascertained solely on the basis of a death certificate, with no other information on the case available after the registry has completed a routine procedure known as “death clearance and followback,” is another measure of the completeness of case ascertainment.

- **No more than 3% of cases are missing information on sex.**

- **No more than 3% of cases are missing information on age.**

- **No more than 5% of cases are missing information on race.**

- **At least 97% of the registry’s records passed a set of single-field and interfield computerized edits.** Computerized edits are computer programs that test the validity and logic of data components. For example, if (a) a patient received a diagnosis of cancer in 1999, (b) the patient’s age was reported as 80 years, and (c) the patient’s year of birth was reported as 1942, a computerized edit could, without human intervention, identify these components as incompatible. The computerized edits applied to the data in this report are incorporated into NAACCR standards (http://www.naaccr.org) and into the EDITS software designed and maintained by CDC (http://www.cdc.gov/cancer/npcr/tools/edits/).

The measurement error for these criteria may vary in select circumstances, following review by CDC and/or NCI.

**References**


Variable Descriptions

Section: Demographic Data Items

Alternate Patient ID Number ...........................................................................17
Address at Diagnosis – State............................................................................18
Address at Diagnosis – County........................................................................19
USCS Standard ................................................................................................20
USCS9915........................................................................................................21
USCS1115........................................................................................................22
USCS9815........................................................................................................23
USCS0615........................................................................................................24
Address at Diagnosis – Census Region ...........................................................25
Race 1...............................................................................................................26
Race 2...............................................................................................................28
Race Recode.....................................................................................................30
Spanish/Hispanic Origin ..................................................................................31
NHIA Derived Hispanic Origin .......................................................................33
IHS Link...........................................................................................................35
State Race Eth Suppress...................................................................................36
Sex....................................................................................................................37
Age at Diagnosis ..............................................................................................38
Age Recode......................................................................................................39
Birth Date.........................................................................................................40
Economic Status...............................................................................................41
Section: Cancer Identification Data Items

Sequence Number – Central .................................................................43
Date of Diagnosis ...............................................................................46
Primary Site .......................................................................................47
Laterality ...........................................................................................48
Grade ..................................................................................................49
Diagnostic Confirmation .................................................................51
Type of Reporting Source ...............................................................52
Histologic Type ICD-O-3 .................................................................53
Behavior Code ICD-O-3 .................................................................54
Behavior Recode for Analysis .......................................................55
Primary Site Recode .......................................................................56
Primary Site Recode with Mesothelioma and Kaposi Sarcoma .......57
SEER International Classification of Childhood Cancer (ICCC) Recode......58

Section: Stage/Prognostic Factors Data Items

SEER Summary Stage 2000 ..............................................................59
SEER Summary Stage 1977 ..............................................................61
CS Extension ...................................................................................63
CS Lymph Nodes .............................................................................64
CS Mets at DX ..................................................................................65
CS Site-Specific Factor 1 ...............................................................66
CS Site-Specific Factor 2 ...............................................................67
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: Demographic Data Items
Alternate Patient ID Number

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random_ID</td>
<td>8</td>
<td>NAACCR Item #20</td>
</tr>
</tbody>
</table>

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.

Considerations for Use
None noted
Section: Demographic Data Items
Address at Diagnosis – State

| 80  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
In addition to USPS abbreviations:
- CD Resident of Canada, NOS (province/territory unknown)
- US Resident of United States, NOS (state/commonwealth/territory/possession unknown)
- XX Resident of country other than the United States (including its territories, commonwealths, or possessions) or Canada, and country is known
- YY Resident of country other than the United States (including its territories, commonwealths, or possessions) or Canada, and country is unknown
- ZZ Residence unknown

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

<table>
<thead>
<tr>
<th>Diagnostic Year(s)</th>
<th>Exclusion state(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>Arkansas, Georgia, New Hampshire, Mississippi, South Dakota, Tennessee, Puerto Rico</td>
</tr>
<tr>
<td>1999</td>
<td>Arkansas, Mississippi, South Dakota, Puerto Rico</td>
</tr>
<tr>
<td>2000</td>
<td>Arkansas, Mississippi, South Dakota, Puerto Rico</td>
</tr>
<tr>
<td>2001</td>
<td>Mississippi, Puerto Rico</td>
</tr>
<tr>
<td>2002</td>
<td>Mississippi, Puerto Rico</td>
</tr>
<tr>
<td>2003</td>
<td>Puerto Rico</td>
</tr>
<tr>
<td>2004</td>
<td>Puerto Rico</td>
</tr>
</tbody>
</table>
Section: Demographic Data Items
Address at Diagnosis – County

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I90_CountyDx</td>
<td>3</td>
<td>NAACCR Item #90</td>
</tr>
</tbody>
</table>

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.

Note: See Appendix A for standard FIPS county codes.

Codes

In addition to FIPS and Geocodes:

- 000 United States, Not Otherwise Specified
- 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).

The following states did not allow permission for their county data to be used: Kansas and Minnesota. The County at Diagnosis variable for these states has been recoded to 000 for all diagnosis years.
**Section: Demographic Data Items**

**USCS Standard**

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>USCSSTD</td>
<td>1</td>
<td>NPCR</td>
</tr>
</tbody>
</table>

**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.

**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, USCS1115 (includes diagnosis years 2011–2015), USCS0615 (includes diagnosis years 2006–2015), USCS9915 (includes diagnosis years 1999-2015) or USCS9815 (includes diagnosis years 1998–2015).
Section: Demographic Data Items
USCS9915

USCS9915

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>USCS9915</td>
<td>1</td>
<td>NPCR</td>
</tr>
</tbody>
</table>

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 1999–2015. When using this variable, restrict the diagnosis years to 1999–2015.

Considerations for use

- This variable is used for analysis of combined 1999–2015 data in the 1998–2015 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, USCS1115 (includes diagnosis years 2011–2015), USCS0615 (includes diagnosis years 2006–2015), USCS9915 (includes diagnosis years 1999-2015) or USCS9815 (includes diagnosis years 1998–2015).
Section: Demographic Data Items
USCS1115

USCS1015

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>USCS1115</td>
<td>1</td>
<td>NPCR</td>
</tr>
</tbody>
</table>

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 2011–2015 (the most recently submitted 5 years of data). When using this variable, restrict the diagnosis years to 2011–2015.

Considerations for use

- This variable is used for analysis of combined 2011–2015 data in the 1998–2015 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, USCS1115 (includes diagnosis years 2011–2015), USCS0615 (includes diagnosis years 2006–2015), USCS9915 (includes diagnosis years 1999-2015) or USCS9815 (includes diagnosis years 1998–2015).
**Section: Demographic Data Items**

**USCS9815**

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>USCS9815</td>
<td>1</td>
<td>NPCR</td>
</tr>
</tbody>
</table>

**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 1998–2015. When using this variable, restrict the diagnosis years to 1998–2015.

**Considerations for use**

- This variable is used for analysis of combined 1998–2015 data in the 1998–2015 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, USCS1115 (includes diagnosis years 2011–2015), USCS0615 (includes diagnosis years 2006–2015), USCS9915 (includes diagnosis years 1999-2015) or USCS9815 (includes diagnosis years 1998–2015).
Section: Demographic Data Items

USCS0615

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>USCS0615</td>
<td>1</td>
<td>NPCR</td>
</tr>
</tbody>
</table>

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 2006–2015 (the most recently submitted 10 years of data). When using this variable, restrict the diagnosis years to 2006–2015.

Considerations for use

- This variable is used for analysis of combined 2006–2015 data in the 1998–2015 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, USCS1115 (includes diagnosis years 2011–2015), USCS0615 (includes diagnosis years 2006–2015), USCS9915 (includes diagnosis years 1999-2015) or USCS9815 (includes diagnosis years 1998–2015).
- The U.S. Cancer Statistics publication standard is available at [https://www.cdc.gov/cancer/npcr/uses/technical_notes/criteria.htm](https://www.cdc.gov/cancer/npcr/uses/technical_notes/criteria.htm).
Section: Demographic Data Items
Address at Diagnosis – Census Region

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENSUS_REGION</td>
<td>9</td>
<td>Derived based upon NAACCR Item #80</td>
</tr>
</tbody>
</table>

**Description**

The region where the patient lived at diagnosis.

**Codes**

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 [http://www.census.gov/geo/www/us_regdiv.pdf](http://www.census.gov/geo/www/us_regdiv.pdf) for a list of states for each region. US Census Regions are Northeast, Midwest, South, and West.

**Considerations for Use**

Do not confuse this variable with census tract. Census tract data are not included in this dataset. Note: CENSUS_REGION is missing for PR.
Section: Demographic Data Items

Race 1

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I160_Race1</td>
<td>2</td>
<td>NAACCR Item #160</td>
</tr>
</tbody>
</table>

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](http://www.census.gov/prod/cen2000/doc/sf2.pdf)

Rationale

Because race has a significant association with cancer rates and outcomes, a comparison between areas with different racial distributions may require an analysis of race to interpret the findings. The race codes listed correspond closely to race categories used by the U.S. Census Bureau to allow calculation of race-specific incidence rates. The full coding system should be used to allow accurate national comparison and collaboration, even if the state population does not include many of the race categories.

Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>White</td>
</tr>
<tr>
<td>02</td>
<td>Black</td>
</tr>
<tr>
<td>03</td>
<td>American Indian, Aleutian, or Eskimo (includes all indigenous populations of the Western hemisphere)</td>
</tr>
<tr>
<td>04</td>
<td>Chinese</td>
</tr>
<tr>
<td>05</td>
<td>Japanese</td>
</tr>
<tr>
<td>06</td>
<td>Filipino</td>
</tr>
<tr>
<td>07</td>
<td>Hawaiian</td>
</tr>
<tr>
<td>08</td>
<td>Korean</td>
</tr>
<tr>
<td>10</td>
<td>Vietnamese</td>
</tr>
<tr>
<td>11</td>
<td>Laotian</td>
</tr>
<tr>
<td>12</td>
<td>Hmong</td>
</tr>
<tr>
<td>13</td>
<td>Kampucheian</td>
</tr>
<tr>
<td>14</td>
<td>Thai</td>
</tr>
<tr>
<td>15</td>
<td>Asian Indian or populations of the Western hemisphere</td>
</tr>
<tr>
<td>20</td>
<td>Micronesian, NOS (code 09 prior to Version 12)</td>
</tr>
<tr>
<td>21</td>
<td>Chamorro/Chamoru</td>
</tr>
<tr>
<td>22</td>
<td>Guamanian, NOS</td>
</tr>
<tr>
<td>09</td>
<td>Asian Indian or Pakistani, NOS (code 09 prior to Version 12)</td>
</tr>
<tr>
<td>26</td>
<td>Hawaiian</td>
</tr>
<tr>
<td>27</td>
<td>Samoan</td>
</tr>
<tr>
<td>28</td>
<td>Tongan</td>
</tr>
<tr>
<td>30</td>
<td>Melanesian, NOS</td>
</tr>
<tr>
<td>31</td>
<td>Fiji Islander</td>
</tr>
<tr>
<td>32</td>
<td>New Guinean</td>
</tr>
<tr>
<td>96</td>
<td>Other Asian, including Micronesian, NOS</td>
</tr>
<tr>
<td>97</td>
<td>Pacific Islander, NOS</td>
</tr>
<tr>
<td>98</td>
<td>Other</td>
</tr>
<tr>
<td>99</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Considerations for Use

Population data are not available for this variable. For age-adjusted rates by race, “NPCR Race Recode” should be used. This variable is used to derive the “NPCR Race Recode”.

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

- Data for American Indians and Alaska Natives cannot be displayed for Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York.
- Data for Asians and Pacific Islanders cannot be displayed for Delaware, Kansas and Kentucky.
- Data for race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
Section: Demographic Data Items
Race 2

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I161_Race2</td>
<td>2</td>
<td>NAACCR Item #161</td>
</tr>
</tbody>
</table>

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

Rationale

Because race has a significant association with cancer rates and outcomes, a comparison between areas with different racial distributions may require an analysis of race to interpret the findings. The race codes listed correspond closely to race categories used by the U.S. Census Bureau to allow calculation of race-specific incidence rates. The full coding system should be used to allow accurate national comparison and collaboration, even if the state population does not include many of the race categories.

Codes

01 White
02 Black
03 American Indian, Aleutian, or Eskimo (includes all indigenous populations of the Western hemisphere)
04 Chinese (prior to Version 12)
05 Japanese
06 Filipino
07 Hawaiian
08 Korean
10 Vietnamese
11 Laotian
12 Hmong
13 Kampuchean
14 Thai
15 Asian Indian or populations of the Western hemisphere
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,
97 Pacific Islander, NOS
98 Other
99 Unknown

Considerations for Use

Population data are not available for this variable. For age-adjusted rates by race, “NPCR Race Recode” should be used. This variable is used to derive the “NPCR Race Recode”.

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

- Data for American Indians and Alaska Natives cannot be displayed for Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York.
• Data for Asians and Pacific Islanders cannot be displayed for Delaware, Kansas and Kentucky.
• Data for race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
Section: Demographic Data Items
Race Recode

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race_Recode_For_USCS</td>
<td>8</td>
<td>Derived based upon NAACCR Items #160, #161, and #192</td>
</tr>
</tbody>
</table>

Description

The variable contains Indian Health Service-linked American Indian data. For this dataset, all NPCR cancer registries linked cancer cases diagnosed from 1995–2015 with the IHS patient registration database to identify American Indians/Alaska Natives that were classified in the registry as non-native. IHS provides medical services to American Indians/Alaska Natives who are members of federally recognized tribes, estimated to be approximately 65% of the American Indian/Alaska Native population. This variable is created from Race1, Race2, and the Indian Health Service (IHS) Link variable. Race/ethnicity starts as Race1. If Race1 is white and Race 2 is a specified non-white race, then the value from Race2 is used. After this check, if Race/ethnicity is still white and there is a positive IHS Link, then Race/Ethnicity is set to American Indian/Alaskan Native.

The “race recode” variable also 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.

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

Codes

This variable combines race into the following categories:
- White
- Black
- American Indian/Alaska Native
- Asian/Pacific Islander
- Other
- Unknown
Section: Demographic Data Items
Spanish/Hispanic Origin

### Description

Code identifying persons of Spanish or Hispanic origin. This code is used by hospital and central registries to show the “best guess” as to whether or not the person should be classified as Hispanic for purposes of calculating cancer rates. If the patient has multiple tumors, all records should have the same code.

Reference to Census 2000 definitions for ethnicity and race:

All information resources should be used to determine the correct code, including:
- Stated ethnicity in the medical record
- Stated Hispanic origin on the death certificate
- Birthplace
- Information about life history and/or language spoken found during the abstracting process
- Patient’s last name [2230] or maiden name [2390] found on a list of Hispanic names

Some registries code the information from the medical record, others code ethnicity based on Spanish names, and others use a combination of methods.

Persons of Spanish or Hispanic origin may be of any race, but these categories generally are not used for Native Americans, Filipinos, etc., who may have Spanish names. If a patient has a Hispanic name, but there is reason to believe they are not Hispanic (e.g., the patient is Filipino, or the patient is a woman known to be non-Hispanic who has a Hispanic married name), the code in this field should be 0 (non-Spanish, non-Hispanic).

Code 7 is assigned if Hispanic ethnicity is based strictly on a computer list or algorithm (unless contrary evidence is available).

**Note:** NAACCR recognizes that available definitions and abstracting instructions for Name--Last [2230] and Name--Maiden [2390] may be inadequate for describing names used in some cultures, including Hispanic cultures. Explicit instructions have not been provided for entering compound names, with or without hyphens or “De.” Order of names, use of maternal and paternal names, and use of hyphens can vary across cultures. It is likely that abstracting and coding practice for these items varies across registries. Limitations inherent in these definitions should be kept in mind when using the data.
Rationale
Ethnic origin has a significant association with cancer rates and outcomes. Hispanic populations have different patterns of occurrence of cancer from other populations that may be included in the “white” category of Race [160].

Codes
0 Non-Spanish; non-Hispanic
1 Mexican (includes Chicano)
2 Puerto Rican
3 Cuban
4 South or Central American (except Brazil)
5 Other specified Spanish/Hispanic origin (includes European; excludes Dominican Republic)
6 Spanish, NOS
   Hispanic, NOS
   Latino, NOS
   There is evidence, other than surname or maiden name, that the person is Hispanic, but he/she cannot be assigned to any of the other categories 1-5.
7 Spanish surname only ((Code 7 is ordinarily for central registry use only, hospital registrars may use code 7 if using a list of Hispanic surnames provided by their central registry; otherwise, code 9 “unknown whether Spanish or not” should be used.)
   The only evidence of the person’s Hispanic origin is the surname or maiden name and there is no contrary evidence that the person is not Hispanic.
8 Dominican Republic
9 Unknown whether Spanish or not

Note: Code 7 was adopted for use effective with 1994 diagnosis and modified December 1994.

Note: Code 8 was added in Standards Volume II Version 10.2, effective January 2005, however, abstractors may assign code 8 to tumors diagnosed prior to 2005.

Considerations for Use
Due to concerns about under-reporting of Hispanics, the NHIA variable [191] was created to identify Hispanics in 2004. Population data are not available for this variable. For age-adjusted rates by ethnicity, the variable “Origin recode NHIA (Hispanic/Non-Hispanic)” should be used.

- The following states have state-level race or ethnicity data presentation restrictions:
  - Data for Hispanic ethnicity cannot be displayed for Delaware, Kentucky, and Massachusetts.
  - Data for race and ethnicity combinations—white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
States are given the right to suppress ethnicity-specific data every submission year.
Section: Demographic Data Items
NHIA Derived Hispanic Origin

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I191_NHIA</td>
<td>1</td>
<td>NAACCR Item #191</td>
</tr>
</tbody>
</table>

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].

Code 7 (Spanish surname only) of the Spanish/Hispanic Origin [190] data item became effective with 1994 diagnoses. For greater detail, please refer to the technical documentation: http://www.naaccr.org/dat#NHIA.

Rationale

Sometimes despite best efforts to obtain complete information directly from the medical record, information is not available and is reported to the cancer registry as a missing data item. With regard to Hispanic ethnicity, some cancer registries have found it necessary to rely on indirect methods to populate this data element. Registries often have significant numbers or proportions of Hispanic populations in their jurisdiction.

Codes

- 0  Non-Hispanic
- 1  Mexican, by birthplace or other specific identifier
- 2  Puerto Rican, by birthplace or other specific identifier
- 3  Cuban, by birthplace or other specific identifier
- 4  South or Central American (except Brazil), by birthplace or other specific identifier
- 5  Other specified Spanish/Hispanic origin (includes European; excludes Dominican Republic), by birthplace or other specific identifier
- 6  Spanish, NOS; Hispanic, NOS; Latino, NOS
- 7  NHIA surname match only
- 8  Dominican Republic
- Blank  Algorithm has not been run

Note: Code 8 was added in Standards Volume II Version 10.2 effective January 2005.

Considerations for Use

Blank values are allowed for states that chose not to include data for NHIA in this file. Data for NPCR registries that are included in U.S. Cancer Statistics use this variable. For age-adjusted
rates by ethnicity, the user-specified variable “Origin recode NHIA (Hispanic/Non-Hispanic)” should be used.

States are given the right to suppress ethnicity-specific data every submission year. The following states have state-level race or ethnicity data presentation restrictions:

- Data for Hispanic ethnicity cannot be displayed for Delaware, Kentucky, and Massachusetts.
- Data for any race and ethnicity combinations—for example, white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
## Section: Demographic Data Items

### IHS Link

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I192_IHS</td>
<td>1</td>
<td>NAACCR Item #192</td>
</tr>
</tbody>
</table>

**Description**

This variable captures the results of the linkage of the registry database with the Indian Health Service patient registration database.

**Rationale**

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**

NPCR registries with one or more IHS Purchase/Referred Care Service Delivery Area (PRCSDA) county are required to link their database with the IHS patient registration database on an annual basis. SEER-only registries (Connecticut, Hawaii, Iowa, New Mexico generally link their database with the IHS database on an annual basis. Those registries not included in the following list may elect to link with IHS annually, but are required to link every five years. Blank values are allowed for states without PRCSDA counties that chose not to not link with IHS annually or chose to not include data for American Indians/Alaskan Natives in this file. Data for NPCR registries that are included in U.S. Cancer Statistics use this variable. Population data are not available for this variable. For age-adjusted rates by race, the variable “NPCR Race Recode” should be used.

<table>
<thead>
<tr>
<th>Alabama</th>
<th>Massachusetts</th>
<th>Oregon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>Michigan</td>
<td>Pennsylvania</td>
</tr>
<tr>
<td>Arizona</td>
<td>Minnesota</td>
<td>Rhode Island</td>
</tr>
<tr>
<td>California</td>
<td>Mississippi</td>
<td>South Carolina</td>
</tr>
<tr>
<td>Colorado</td>
<td>Montana</td>
<td>South Dakota</td>
</tr>
<tr>
<td>Florida</td>
<td>Nebraska</td>
<td>Texas</td>
</tr>
<tr>
<td>Idaho</td>
<td>Nevada</td>
<td>Utah</td>
</tr>
<tr>
<td>Indiana</td>
<td>New York</td>
<td>Washington</td>
</tr>
<tr>
<td>Kansas</td>
<td>North Carolina</td>
<td>Wisconsin</td>
</tr>
<tr>
<td>Louisiana</td>
<td>North Dakota</td>
<td>Wyoming</td>
</tr>
<tr>
<td>Maine</td>
<td>Oklahoma</td>
<td></td>
</tr>
</tbody>
</table>
Section: Demographic Data Items  
State race eth suppress

State race eth suppress

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stateraceethincl</td>
<td>1</td>
<td>Derived based on NAACCR Item #192</td>
</tr>
</tbody>
</table>

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

Considerations for use
- States have the option to suppress race-specific and Hispanic ethnicity–specific data every submission year. While these states can be included in an aggregated analysis, the affected state’s race and ethnicity information cannot be reported at the state level.
- The following states have state-level race or ethnicity data presentation restrictions:
  - Data for American Indians and Alaska Natives cannot be displayed for Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York.
  - Data for Asians and Pacific Islanders cannot be displayed for Delaware, Kansas, and Kentucky.
  - Hispanic ethnicity data cannot be displayed for Delaware, Kentucky, and Massachusetts.
  - Any race and ethnicity combinations—for example, white Hispanic, white non-Hispanic, black Hispanic, and black non-Hispanic—cannot be displayed for Delaware, Kansas, Kentucky, Massachusetts, and Pennsylvania.
- For more information, please refer to the “Race Recode for USCS” 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
- 3 Other (hermaphrodite)
- 4 Transsexual
- 9 Not stated/Unknown

Considerations for Use
- To get the correct population denominator, “female” must be selected when analyzing female-specific cancers (such as ovarian cancer or female breast cancer) and “male” for male-specific cancers (such as prostate cancer).
- Due to small case counts and the lack of an associated population file, cases for sex other than male or female are excluded from this database.
Section: Demographic Data Items
Age at Diagnosis

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I230 AgeDx</td>
<td>3</td>
<td>NAACCR Item #230</td>
</tr>
</tbody>
</table>

Description
Age of the patient at diagnosis in complete years.

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 should only be used for special analyses; e.g. modeling.
Section: Demographic Data Items
Age Recode

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgeRec</td>
<td>2</td>
<td>Derived based upon NAACCR Item #230</td>
</tr>
</tbody>
</table>

Description
A standard grouping of age at diagnosis into 19 categories (<1, 1-4, 5-9, …75-79, 80-84, 85+).

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

Considerations for Use
None noted.
Section: Demographic Data Items
Birth Date

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I240_DOB</td>
<td>4</td>
<td>Derived based upon NAACCR Item #240</td>
</tr>
</tbody>
</table>

Description
Year of birth of the patient.

Considerations for Use
The day and month of birth are not provided for confidentiality reasons; if age is over 99, then year of birth is recoded.
Section: Demographic Data Items
Economic Status

<table>
<thead>
<tr>
<th>Economic Status</th>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Econ_Status</td>
<td>4</td>
<td>Derive based upon NAACCR Item #90</td>
<td></td>
</tr>
</tbody>
</table>

Description
County level economic status variable as assigned by the Appalachian Regional Commission.

Considerations for Use

- You MUST 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.

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

- **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: https://www.arc.gov/research/sourceandmethodologycountyeconomicstatusfy2007fy2016.asp.
Section: Cancer Identification Data Items
Sequence Number – Central

<table>
<thead>
<tr>
<th>Sequence Number – Central</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I380_SeqNoCntrl</td>
<td>2</td>
<td>NAACCR Item #380</td>
</tr>
</tbody>
</table>

**Description**
Code indicates the sequence of all reportable neoplasms over the lifetime of the person. Each primary neoplasm (not progression or recurrences) is assigned a different number. Sequence Number 00 indicates that the person has had only one in situ or one malignant neoplasm as defined by the Federal reportable list (regardless of central registry reference date). Sequence Number 01 indicates the first of two or more reportable neoplasms, 02 indicates the second of two or more reportable neoplasms, and so on. 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 or was diagnosed and treated in another state.

**Reporting Requirements: Federally Required and State/Province Defined**
The Federal standard defining the reportable neoplasms is described in the NAACCR Volume II Data Standards and Data Dictionary Chapter III, Standards For Tumor Inclusion and Reportability. It is assumed that this shared standard is the “minimum” definition of reportability. Individual central cancer registries may define additional neoplasms as reportable. Numeric codes in the 00-59 range indicate the sequence of neoplasms of in situ or malignant behavior (2 or 3) at the time of diagnosis, which NPCR standards require to be reported. Codes 60 to 87 indicate the sequence of non-malignant tumors (as defined in Chapter III) and any other neoplasms that the central registry has defined as reportable. Neoplasms required by NPCR with an in situ or malignant behavior at the time of diagnosis are sequenced completely independently of this higher-numbered category.

**Rationale**
The purpose of sequencing based on the patient’s lifetime is to truly identify the 00s, the people who only had one malignant primary in their lifetime for survival analysis. If a central registry sequences by just what is reported to them, then it will be unclear whether 00 means the person only had one malignant primary in his lifetime or the person had one malignant primary since the central registry started collecting data. The Federally required reportable list has changed throughout the years, so the registry must use the appropriate reportable list for the year of diagnosis. The central registry reference date will not affect Sequence Number-Central.

**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 \textit{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


The table that follows shows which sequence number series to use by type of neoplasm.

<table>
<thead>
<tr>
<th>Neoplasm</th>
<th>SeqNum-Central (Numeric Series)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{In Situ}/Malignant as Federally Required based on Diagnosis Year</td>
<td></td>
</tr>
<tr>
<td>\textit{In Situ} (behavior code = 2) (Cervix CIS/CIN III, Diagnosis Year before 1996) (includes VIN III, VAIN III, AIN III)</td>
<td>00 -- 59</td>
</tr>
<tr>
<td>Malignant (behavior code = 3)</td>
<td>00 -- 59</td>
</tr>
<tr>
<td>Juvenile Astrocytoma, Diagnosis Year 2001+ (*)</td>
<td>00 -- 59</td>
</tr>
<tr>
<td>Invasive following \textit{In Situ}--New primary as defined by CoC</td>
<td>00 -- 59</td>
</tr>
<tr>
<td>Invasive following \textit{In Situ}--New primary as defined by SEER</td>
<td>00 -- 59</td>
</tr>
<tr>
<td>Unspecified Federally Required Sequence Number or Unknown</td>
<td>99</td>
</tr>
</tbody>
</table>

| 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.
**Section: Cancer Identification Data Items**

**Date of Diagnosis**

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I390 DateDX</td>
<td>6</td>
<td>Derived based upon NAACCR Item #390</td>
</tr>
</tbody>
</table>

**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 1998-2015.

**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
- Blank – when no known date applies
Section: Cancer Identification Data Items
Primary Site

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I400_Site</td>
<td>4</td>
<td>NAACCR Item #400</td>
</tr>
</tbody>
</table>

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

Considerations for Use

See ICD-O-3,\(^{14}\) or ICD-O-3,\(^{13}\) Topography Section, for the codes for primary site.

Consider reviewing the variables “Primary Site Recode” or “Primary Site Recode with Kaposi Sarcoma and Mesothelioma” before using the directly coded primary site. For more information on the SEER primary site recodes, see [http://seer.cancer.gov/siterecode/](http://seer.cancer.gov/siterecode/).
Section: Cancer Identification Data Items
Laterality

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I410 Laterality</td>
<td>1</td>
<td>NAACCR Item #410</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I440 Grade</td>
<td>1</td>
<td>NAACCR Item #440</td>
</tr>
</tbody>
</table>

**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**


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, users should review the site-specific modules available at: http://training.seer.cancer.gov/modules_site_spec.html and the most current FORDS manual (http://www.facs.org/cancer/coc/fordsmanual.html). Each module has an abstracting, coding, and staging section, which has a morphology and grading 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:


Section: Cancer Identification Data Items

Diagnostic Confirmation

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I490_DxConf</td>
<td>1</td>
<td>NAACCR Item #490</td>
</tr>
</tbody>
</table>

Description

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

Rationale

Diagnostic confirmation is useful to calculate rates based on microscopically confirmed cancers. Full incidence calculations must also include tumors that are only confirmed clinically. The percentage of tumors that are not microscopically confirmed is an indication of whether case finding is including sources outside of pathology reports.

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

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I500_TypeRptSrc</td>
<td>1</td>
<td>NAACCR Item #500</td>
</tr>
</tbody>
</table>

### 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).

### Rationale

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.

### 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.

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

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I522_HistTypeICDO3</td>
<td>4</td>
<td>NAACCR Item #522</td>
</tr>
</tbody>
</table>

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.

Considerations for Use

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

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I523_BehavICDO3</td>
<td>1</td>
<td>NAACCR Item #523</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Benign</td>
</tr>
<tr>
<td>1</td>
<td>Uncertain whether benign or malignant</td>
</tr>
<tr>
<td></td>
<td>Borderline malignancy</td>
</tr>
<tr>
<td></td>
<td>Low malignant potential</td>
</tr>
<tr>
<td></td>
<td>Uncertain malignant potential</td>
</tr>
<tr>
<td>2</td>
<td>Carcinoma in situ</td>
</tr>
<tr>
<td></td>
<td>Intraepithelial</td>
</tr>
<tr>
<td></td>
<td>Noninfiltrating</td>
</tr>
<tr>
<td></td>
<td>Noninvasive</td>
</tr>
<tr>
<td>3</td>
<td>Malignant, primary site</td>
</tr>
<tr>
<td>6</td>
<td>Malignant, metastatic site</td>
</tr>
<tr>
<td></td>
<td>Malignant, secondary site</td>
</tr>
<tr>
<td>9</td>
<td>Malignant, uncertain whether primary or metastatic site</td>
</tr>
</tbody>
</table>

Considerations for Use
Section: Cancer Identification Data Items
Behavior Recode for Analysis

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior_Recode_For_Analysis</td>
<td>8</td>
<td>Derived based upon NAACCR Items #400, #522, and #523</td>
</tr>
</tbody>
</table>

**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.

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

**Codes**


See Appendix II for re-code details.

Note: Behavior_Recode_For_Analysis is missing for PR.
Section: Cancer Identification Data Items
Primary Site Recode

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site_Recode_ICD_O_3_WH</td>
<td>8</td>
<td>Derived based upon NAACCR Items #400 and #522</td>
</tr>
</tbody>
</table>

**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 re-code details.

Section: Cancer Identification Data Items
Primary Site Recode
with Mesothelioma and Kaposi Sarcoma

Primary Site Recode with Mesothelioma and Kaposi Sarcoma

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiterecwithKaposiandmesothelio</td>
<td>8</td>
<td>Derived based upon NAACCR Items #400 and #522</td>
</tr>
</tbody>
</table>

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 pulls out Mesothelioma and Kaposi Sarcoma as separate categories and is the recode used by CDC, NCI, and NAACCR in their surveillance publications. This recode is defined by the SEER program.

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

Codes

See Appendix II for re-code details.

Note: SEER reference for Primary Site Recode for ICD-O-3 is:

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

<table>
<thead>
<tr>
<th>SEER-Modified International Classification of Childhood Cancer (ICCC) Recode</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICCCSiteRecExtendedICDO3WHO200</td>
</tr>
</tbody>
</table>

**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 re-code details.

Note: beginning with data released in 2006, the grouping of childhood cancers is based on ICD-O-3 instead of ICD-O-2. SEER reference for the ICCC recodes is: [http://seer.cancer.gov/iccc/](http://seer.cancer.gov/iccc/)

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

SEER Summary Stage 2000

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I759_SS2000</td>
<td>1</td>
<td>NAACCR Item #759</td>
</tr>
</tbody>
</table>

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.

Rationale
Stage information is important when evaluating the effects of cancer control programs. It is crucial in understanding whether changes over time in incidence rates or outcomes are due to earlier detection of the cancers. In addition, cancer treatment cannot be studied without knowing the stage at diagnosis.

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. Summary Stage 2000 is used for tumors diagnosed 2001-2003.

For cases diagnosed 2004-2014, Summary Stage 2000 is derived from information coded in the Collaborative Stage Data Collection System. In certain circumstances when information is not
available to calculate the Derived Summary Stage 2000, NPCR permits the use of Summary Stage 2000 for diagnosis years 2004-2014. Summary Stage 2000 or Derived Summary Stage 2000 could be used for cases diagnosed in 2015. As a result, when stage information is required, it is recommended that and the “Merged Summary Stage 2000” be used to account for missing information in Derived Summary Stage 2000.

Cases diagnosed before January 1, 2001, are assigned a summary stage according to SEER Summary Stage Guide 1977, and the code is reported in SEER Summary Stage 1977 [760].

To assess the effect of changes between Summary Stage 2000 and Summary Stage 1977 on a particular site, there are two references: 1) Phillips, JL, coordinator. Data Effects of the Changes in 2000 and 2) Summary Stage Comparability Report, 2005 from the Collaborative Research Working Group of NAACCR. Both are located on the NAACCR web site at www.naaccr.org.

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

<table>
<thead>
<tr>
<th>Diagnosis Years</th>
<th>Summary Stage Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998-2000</td>
<td>Summary Stage 1977</td>
</tr>
<tr>
<td>2001-2003</td>
<td>Summary Stage 2000</td>
</tr>
<tr>
<td>2004-2014</td>
<td>Derived Summary Stage 2000 (see note above)</td>
</tr>
<tr>
<td>2015</td>
<td>Summary Stage 2000 or Derived Summary Stage 2000 (see note above)</td>
</tr>
</tbody>
</table>

See Cautionary Notes – Stage for additional information.
Section: Stage/Prognostic Factors Data Items
SEER Summary Stage 1977

SEER Summary Stage 1977

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I760 SS1977</td>
<td>1</td>
<td>NAACCR Item #760</td>
</tr>
</tbody>
</table>

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.

Rationale
Stage information is important when evaluating the effects of cancer control programs. It is crucial for understanding whether changes over time in incidence rates or outcomes are due to earlier detection of the cancers. In addition, cancer treatment cannot be studied without knowing the stage at diagnosis.

To study historical trends in stage, the coding system must be relatively unchanged (stable) over time. AJCC’s TNM system is updated periodically to maintain clinical relevance with changes in diagnosis and treatment. The surveillance registries often rely on the Summary Stage, which they consider to be more “stable.” Summary Stage has been in widespread use, either as the primary staging scheme or a secondary scheme, in most central and hospital registries since 1977.

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. Summary Stage 2000 is used for tumors diagnosed 2001-2003.

For cases diagnosed 2004-2014, Summary Stage 2000 is derived from information coded in the Collaborative Stage Data Collection System. In certain circumstances when information is not available to calculate the Derived Summary Stage 2000, NPCR permits the use of Summary Stage 2000 for diagnosis years 2004-2014. Summary Stage 2000 or Derived Summary Stage 2000 could be used for cases diagnosed in 2015. As a result, when stage information is required, it is recommended that the “Merged Summary Stage 2000” variable be used to account for missing information in Derived Summary Stage 2000.

Cases diagnosed before January 1, 2001, are assigned a summary stage according to SEER Summary Stage Guide 1977, and the code is reported in SEER Summary Stage 1977 [760].

To assess the effect of changes between Summary Stage 2000 and Summary Stage 1977 on a particular site, there are two references: 1) Phillips, JL, coordinator. Data Effects of the Changes in 2000 and 2) Summary Stage Comparability Report, 2005 from the Collaborative Research Working Group of NAACCR. Both are located on the NAACCR web site at www.naaccr.org.

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

<table>
<thead>
<tr>
<th>Diagnosis Years</th>
<th>Summary Stage Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998-2000</td>
<td>Summary Stage 1977</td>
</tr>
<tr>
<td>2001-2003</td>
<td>Summary Stage 2000</td>
</tr>
<tr>
<td>2004-2014</td>
<td>Derived Summary Stage 2000 (see note above)</td>
</tr>
<tr>
<td>2015</td>
<td>Summary Stage 2000 or Derived Summary Stage 2000 (see note above)</td>
</tr>
</tbody>
</table>

See Cautionary Notes – Stage for additional information.
Section: Stage/Prognostic Factors Data Items
CS Extension

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2810_CSExt</td>
<td>3</td>
<td>NAACCR Item #2810</td>
</tr>
</tbody>
</table>

**Description**

This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual*, 6th and 7th editions, as well as the SEER Summary Stage, 1977 and 2000. CS Extension identifies primary tumor growth (extension) within the organ of origin or its extension into neighboring organs. For certain sites, such as ovary, discontinuous metastasis is coded in *CS Extension*. Site-specific codes provide extensive detail describing disease extent.

**Rationale**

Tumor extension at diagnosis is a prognostic indicator used by Collaborative Staging to derive SEER Summary Stage codes.

**Codes**

(See the most current version of the *Collaborative Stage Data Collection System* ([http://cancerstaging.org](http://cancerstaging.org)) for rules and site-specific codes and coding structures.)

**Considerations for Use**

This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.
Section: Stage/Prognostic Factors Data Items
CS Lymph Nodes

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2830_CSLymphNodes</td>
<td>3</td>
<td>NAACCR Item #2830</td>
</tr>
</tbody>
</table>

Description

This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual*, 6th and 7th editions, as well as the SEER Summary Stage, 1977 and 2000. CS Lymph Nodes is site-specific and identifies the regional lymph nodes involved with cancer at the time of diagnosis.

Rationale

The involvement of specific regional lymph nodes is a prognostic indicator used by Collaborative Staging to derive SEER Summary Stage codes.

Codes

(See the most current version of the Collaborative Stage Data Collection System ([http://cancerstaging.org](http://cancerstaging.org)) for rules and site-specific codes and coding structures.)

Considerations for Use

This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.
Section: Stage/Prognostic Factors Data Items

CS Mets at DX

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2850_CSMetsDx</td>
<td>3</td>
<td>NAACCR Item #2850</td>
</tr>
</tbody>
</table>

Description
This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual, 6th and 7th editions*, as well as the SEER Summary Stage, 1977 and 2000. CS Lymph Nodes is site-specific and identifies the site(s) of metastatic involvement at the time of diagnosis.

Rationale
The presence of metastatic disease at diagnosis is an independent prognostic indicator used by Collaborative Staging to derive SEER Summary Stage codes.

Codes
(See the most current version of the Collaborative Stage Data Collection System [http://cancerstaging.org](http://cancerstaging.org) for rules and site-specific codes and coding structures.)

Considerations for Use
This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.
Section: Stage/Prognostic Factors Data Items
CS Site Specific Factor 1

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2880_CSSSF1</td>
<td>3</td>
<td>NAACCR Item #2880</td>
</tr>
</tbody>
</table>

Description
This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual, 6th* and *7th* editions, as well as the SEER Summary Stage, 1977 and 2000. CS Site Specific Fact 1 (SSF1) identifies additional information needed to generate stage, or prognostic factors that have an effect on stage or survival.

Rationale
Site-specific factors are used to record additional staging information needed by Collaborative Staging to derive SEER Summary Stage codes for particular site-histology schema.

Codes
The information recorded in CS Site-Specific Factor 1 differs for each anatomic site. See the most current version of the Collaborative Stage Data Collection System ([http://cancerstaging.org](http://cancerstaging.org)) for rules and site-specific codes and coding structures.

Considerations for Use
This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.

NPCR requires CS SSF1 for the female breast, brain/CNS, lung, pleura, and retinoblastoma primary sites/histologies only. SEER requires CS SSF1 for several primary sites/histologies. Though CS SSF1 may be included in the dataset for other primary sites/histologies, the CS SSF1 information for those primary sites/histologies reported by NPCR central cancer registries was reported as it was available and is considered unreliable at this time.
Section: Stage/Prognostic Factors Data Items
CS Site Specific Factor 2

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2890_CSSSF2</td>
<td>3</td>
<td>NAACCR Item #2890</td>
</tr>
</tbody>
</table>

**Description**

This data items belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual, 6th* and *7th* editions, as well as the SEER Summary Stage, 1977 and 2000. CS Site Specific Fact 2 (SSF2) identifies additional information needed to generate stage, or prognostic factors that have an effect on stage or survival.

**Rationale**

Site-specific factors are used to record additional staging information needed by Collaborative Staging to derive SEER Summary Stage codes for particular site-histology schema.

**Codes**

*(The information recorded in CS Site-Specific Factor 2 differs for each anatomic site. See the most current version of the Collaborative Stage Data Collection System (http://cancerstaging.org) for rules and site-specific codes and coding structures.)*

**Considerations for Use**

This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.

NPCR requires CS SSF2 for the female breast and corpus uteri primary sites/histologies only. SEER requires CS SSF1 for several primary sites/histologies. Though CS SSF1 may be included in the dataset for other primary sites/histologies, the CS SSF1 information for those primary sites/histologies reported by NPCR central cancer registries was reported as it was available and is considered unreliable at this time.
Section: Stage/Prognostic Factors Data Items
CS Site Specific Factor 3

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2900_CSSSF3</td>
<td>3</td>
<td>NAACCR Item #2900</td>
</tr>
</tbody>
</table>

Description
This data items belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual*, 6th and 7th editions, as well as the SEER Summary Stage, 1977 and 2000. CS Site Specific Fact 3 (SSF3) identifies additional information needed to generate stage, or prognostic factors that have an effect on stage or survival.

Rationale
Site-specific factors are used to record additional staging information needed by Collaborative Staging to derive SEER Summary Stage codes for particular site-histology schema.

Codes
(The information recorded in CS Site-Specific Factor 3 differs for each anatomic site. See the most current version of the Collaborative Stage Data Collection System [(http://cancerstaging.org)](http://cancerstaging.org) for rules and site-specific codes and coding structures.)

Considerations for Use
This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.

NPCR requires CS SSF3 for the prostate primary site/histologies only. SEER requires CS SSF1 for several primary sites/histologies. Though CS SSF1 may be included in the dataset for other primary sites/histologies, the CS SSF1 information for those primary sites/histologies reported by NPCR central cancer registries was reported as it was available and is considered unreliable at this time.
Section: Stage/Prognostic Factors Data Items
CS Site Specific Factor 15

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2869_CSSSF15</td>
<td>3</td>
<td>NAACCR Item #2869</td>
</tr>
</tbody>
</table>

Description
This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual*, 6th and 7th editions, as well as the SEER Summary Stage, 1977 and 2000. CS Site Specific Fact 15 (SSF15) identifies additional information needed to generate stage, or prognostic factors that have an effect on stage or survival.

Rationale
Site-specific factors are used to record additional staging information needed by Collaborative Staging to derive SEER Summary Stage codes for particular site-histology schema.

Codes
(The information recorded in CS Site-Specific Factor 15 differs for each anatomic site. See the most current version of the Collaborative Stage Data Collection System ([http://cancerstaging.org](http://cancerstaging.org)) for rules and site-specific codes and coding structures.)

Considerations for Use
This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.

NPCR requires CS SSF15 for the female breast primary site/histologies only. SEER requires CS SSF1 for several primary sites/histologies. Though CS SSF1 may be included in the dataset for other primary sites/histologies, the CS SSF1 information for those primary sites/histologies reported by NPCR central cancer registries was reported as it was available and is considered unreliable at this time.
Section: Stage/Prognostic Factors Data Items
CS Site Specific Factor 25

<table>
<thead>
<tr>
<th>CS Site Specific Factor 25</th>
<th>Description</th>
<th>Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2879_CSSSF25</td>
<td>NAACCR Item #2879; AJCC</td>
<td>(The information recorded in CS Site-Specific Factor25 differs for each anatomic site. See the most current version of the Collaborative Stage Data Collection System (<a href="http://cancerstaging.org">http://cancerstaging.org</a>) for rules and site-specific codes and coding structures.)</td>
</tr>
</tbody>
</table>

Description
This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual*, 6th and 7th editions, as well as the SEER Summary Stage, 1977 and 2000. CS Site Specific Fact 25 (SSF25) identifies additional information needed to generate stage.

Rationale
CS Site-Specific Factor25 is used to discriminate between CS staging schema where site and histology alone are insufficient to identify the tumor type or location to identify the applicable staging method.

Considerations for Use
This variable provides detailed information and allows stratification that may not be available when using the stage grouping. Caution should be used to avoid stratification that may introduce biases due to small case counts.

CS SSF25 is required for the following primary sites only:

- BileDuctsDistal
- BileDuctsPerihilar
- CysticDuct
- EsophagusGEJunction
- LacrimalGland
- LacrimalSac
- MelanomaCiliaryBody
- MelanomaIris
- Nasopharynx
- Peritoneum
- PeritoneumFemaleGen
- PharyngealTonsil
- Stomach
### Section: Stage/Prognostic Factors Data Items

#### CS Version Input Original

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2935_CSVerInputOrig</td>
<td>6</td>
<td>NAACCR Item #2935; AJCC</td>
</tr>
</tbody>
</table>

#### Description

This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual*, 6th and 7th editions, as well as the SEER Summary Stage, 1977 and 2000. CS Version Input Original indicates the number of the version initially used to code Collaborative Staging (CS) fields. The CS version number is returned as part of the output of the CS algorithm.

#### Rationale

Over time, the input codes and instructions for CS items may change. This item identifies the correct interpretation of input CS items.

#### Codes

(See the most current version of the Collaborative Stage Data Collection System ([http://cancerstaging.org](http://cancerstaging.org)) for rules and site-specific codes and coding structures.)

CS Version Input Original is a 6-digit code (e.g., 010100). The first two digits represent the major version number; the second two digits represent minor version changes; and, the last two digits represent even less significant changes, such as corrections of typographical errors that do not affect coding or derivation of results.

This item should not be blank if the CS Derived items contain values. It should be blank if the CS Derived items are empty or the CS algorithm has not been applied.

#### Considerations for Use

None noted
Section: Stage/Prognostic Factors Data Items
CS Version Derived

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2936_CSVerDerived</td>
<td>6</td>
<td>NAACCR Item #2936; AJCC</td>
</tr>
</tbody>
</table>

Description
This data items belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the *AJCC Cancer Staging Manual, 6th* and *7th* editions, as well as the SEER Summary Stage, 1977 and 2000. CS Version Input Derived indicates the used to derive Collaborative Staging (CS) fields and is recorded the first time the CS output fields are derived and are updated each time the CS Derived items are recomputed. The CS version number is returned as part of the output of the CS algorithm.

Rationale
The CS algorithm may be re-applied to compute the CS Derived items; for example, when the data are to be used for a special study, transmitted, or when an updated CS algorithm is produced. This item identifies the specific algorithm used to obtain the CS Derived values in the data record.

Codes
*(See the most current version of the Collaborative Stage Data Collection System (http://cancerstaging.org) for rules and site-specific codes and coding structures.)*

CS Version Derived is a 6-digit code (e.g., 010100). The first two digits represent the major version number; the second two digits represent minor version changes; and, the last two digits represent even less significant changes, such as corrections of typographical errors that do not affect coding or derivation results.

This item should not be blank if the CS Derived items contain values. It should be blank if the CS Derived items are empty or the CS algorithm has not been applied.

Considerations for Use
None noted
Section: Stage/Prognostic Factors Data Items
CS Input Current

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2937_CSVerInputCur</td>
<td>6</td>
<td>NAACCR Item #2937; AJCC</td>
</tr>
</tbody>
</table>

Description
This data item belongs to the Collaborative Stage (CS) Data Collection System. The CS is based on the AJCC Cancer Staging Manual, 6th and 7th editions, as well as the SEER Summary Stage, 1977 and 2000. CS Input Current identifies the version used to code Collaborative Staging (CS) fields after they have been updated or recoded. This data item is recorded the first time the CS input fields are entered and should be updated each time the CS input fields are modified.

Rationale
Over time, the input codes and instructions for CS items may change. This item identifies the correct interpretation of input CS items.

Codes
(See the most current version of the Collaborative Stage Data Collection System (http://cancerstaging.org) for rules and site-specific codes and coding structures.)

CS Version Input Current is a 6-digit code (e.g., 020100). The first two digits represent the major version number; the second two digits represent minor version changes; and, the last two digits represent even less significant changes, such as corrections of typographical errors that do not affect coding or derivation of results.

This item should not be blank if the CS Derived items contain values. It should be blank if the CS Derived items are empty or the CS algorithm has not been applied.

Considerations for Use
None noted
Section: Stage/Prognostic Factors Data Items
Derived SEER Summary State 2000

**Derived SS2000**

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I3020_DerivedSS2000</td>
<td>1</td>
<td>NAACCR Item #3020; AJCC</td>
</tr>
</tbody>
</table>

**Description**

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

**Rationale**

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**

(See the most current version of the Collaborative Stage Data Collection System (http://cancerstaging.org/cstage/manuals.html) for rules and site-specific codes and coding structures.)

**Considerations for Use**

Records in this dataset should have a Derived SS2000 for diagnosis years 2004-2015. This data item is usually blank for records in this dataset with a diagnosis year prior to 2004 (1998-2003). In certain circumstances when information is not available to calculate the Derived Summary Stage 2000, NPCR permits the use of Summary Stage 2000 for diagnosis years 2004-2014. Summary Stage 2000 or Derived Summary Stage 2000 could be used for cases diagnosed in 2015. As a result, when stage information is required, it is recommended that consideration be given to using both Derived Summary Stage 2000 and Summary Stage 2000 to account for missing information in Derived Summary Stage 2000. 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:

<table>
<thead>
<tr>
<th>Diagnosis Years</th>
<th>Summary Stage Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998-2000</td>
<td>Summary Stage 1977</td>
</tr>
<tr>
<td>2001-2003</td>
<td>Summary Stage 2000</td>
</tr>
<tr>
<td>2004-2014</td>
<td>Derived Summary Stage 2000 (see note above)</td>
</tr>
<tr>
<td>2015</td>
<td>Summary Stage 2000 or Derived Summary Stage 2000 (see note above)</td>
</tr>
</tbody>
</table>

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. However, particular attention should be paid to data query results and stage information should be used with caution. If there are concerns about stage distributions resulting from data queries, please contact CDC (uscsdata@cdc.gov).

See Cautionary Notes – Stage 3 for additional information.
Section: Stage/Prognostic Factors Data Items
Merged Summary Stage 2000

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merged_SumStage2000</td>
<td>1</td>
<td>NPCR, combined from NAACCR Items #759 and #3020</td>
</tr>
</tbody>
</table>

Description

This is a merged stage variable created using two other variables: “SEER Summary Stage 2000,” which records stage from diagnosis years 2001–2003, “Derived SS2000,” which records stage from diagnostic years 2004–2014, and either variable for diagnostic year 2015. This stage variable can be used for diagnosis years 2001–2015.

Considerations for use

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

See Appendix II for re-code details.
Section: Treatment—First Course
RX Summ—Surgery Primary Site

RX Summ—Surgery Primary Site

| 11290_RxSummSurgPrimSite |   1 | 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.

Considerations for use

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

- 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 FORDS and SEER Program Code manual for additional instructions:

Page 77 of 114
Section: Treatment—First Course
Merged Radiation

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merged_Radiation</td>
<td>1</td>
<td>Derived from NAACCR Items #1360 and #1570</td>
</tr>
</tbody>
</table>

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

Considerations for use

Codes

1. Had radiation
2. Did not have radiation
3. Patient or guardian refused radiation
4. Radiation recommended but unknown if received

- This variable is only available for female breast and for cases submitted by NPCR central cancer registries. The data are not available for the following SEER-only states: Connecticut, Hawaii, Iowa, and New Mexico.
- Data for this variable are available starting with **diagnosis year 2003**.

See Appendix II for re-code details.
Section: Over-ride Flags Data Items
Over-ride Age/Site/Morph

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1990_ORAgeSiteMorph</td>
<td>1</td>
<td>NAACCR Item #1990; SEER</td>
</tr>
</tbody>
</table>

**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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Primary Site, Morphology ICDO3 (SEER IF15)</td>
<td>Identifies records with an unusual occurrence of a particular age/site/histology combination for a given age group</td>
</tr>
<tr>
<td>Age, Primary Site, Morph ICDO3--Adult (SEER)</td>
<td>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 &gt;15</td>
</tr>
<tr>
<td>Age, Primary Site, Morph ICDO3--Pediatric (NPCR)</td>
<td>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</td>
</tr>
</tbody>
</table>

**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 cases 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)

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2000_ORSeqNoDxConf</td>
<td>1</td>
<td>NAACCR Item #2000; SEER</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Confirm, Seq Num--Central (SEER IF23)</td>
<td>Identifies records with multiple primary cancers where at least one primary cancer is not microscopically confirmed</td>
</tr>
</tbody>
</table>

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 Inter-record Review (Inter-record Edit 09)

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2010_ORSiteLatSeqNo</td>
<td>1</td>
<td>NAACCR Item #2010; SEER</td>
</tr>
</tbody>
</table>

**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 must be 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:

<table>
<thead>
<tr>
<th>Inter-record Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verify Same Primary Not Reported Twice for a Person</td>
<td>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</td>
</tr>
<tr>
<td>(SEER IR09)</td>
<td></td>
</tr>
</tbody>
</table>

**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**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reviewed and confirmed as reported</td>
</tr>
<tr>
<td>Blank</td>
<td>Not reviewed or reviewed and corrected</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2030_ORSiteType</td>
<td>1</td>
<td>NAACCR Item #2030; SEER</td>
</tr>
</tbody>
</table>

**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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Site, Morphology-Type, Behavior ICDO3 (SEER IF25)</td>
<td>Identifies records where the site/histology/behavior combination is not in the SEER Site/Histology Validation List</td>
</tr>
</tbody>
</table>

**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/](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

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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Confirmation, Behavior ICDO3 (SEER IF31)</td>
<td>Identifies records with a behavior of in situ and a non-microscopic diagnostic confirmation</td>
</tr>
<tr>
<td>Morphology--Type/Behavior ICDO3 (SEER MORPH)</td>
<td></td>
</tr>
</tbody>
</table>

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. If 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 MOPRH).
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)

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2050 ORRptSrc</td>
<td>1</td>
<td>NAACCR Item #2050; SEER</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Rep Srce(DC), Seq Num—Cent, ICDO3 (SEER IF04)</td>
<td>Identifies records with multiple primary cancers where one is reported only through a death certificate and histology code is &lt;9590</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2060_.ORIlldefineSite</td>
<td>1</td>
<td>NAACCR Item #2060; SEER</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq Num--Central, Prim Site, Morph ICDO3 (SEER IF22)</td>
<td>Identifies records with multiple primary cancers where one is reported as an ill-defined primary site</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>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</td>
</tr>
<tr>
<td>Blank</td>
<td>Not reviewed or reviewed and corrected</td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2070 ORLeukLymph</td>
<td>1</td>
<td>NAACCR Item #2070; SEER</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Confirmation, Histology ICDO3 (SEER IF48)</td>
<td>Identifies leukemia and lymphoma records where the diagnostic confirmation is not microscopic</td>
</tr>
</tbody>
</table>

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. The 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)

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2071_ORSiteBehav</td>
<td>1</td>
<td>NAACCR Item #2071; SEER</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Site, Behavior Code ICDO3 (SEER IF39)</td>
<td>Identifies records with a non-specific primary cancer site code with an <em>in situ</em> behavior</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reviewed and confirmed as reported</td>
</tr>
<tr>
<td>Blank</td>
<td>Not reviewed or reviewed and corrected</td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>SAS Alternate Name</th>
<th>Length</th>
<th>Source of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2074_ORSiteLatMorph</td>
<td>1</td>
<td>NAACCR Item #2074; SEER</td>
</tr>
</tbody>
</table>

### 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:

<table>
<thead>
<tr>
<th>Inter-field Edit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laterality, Primary Site, Morph ICDO3 (SEER IF42)</td>
<td>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.</td>
</tr>
</tbody>
</table>

### 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

1. The data item is left blank if the program does not generate an error message for the edit, Primary site, Morph ICDO3 (SEER IF42) 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 laterality are correctly coded.

### Codes

1. Reviewed and confirmed as reported
2. 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.
Appendix I

NPCR_9915_AnalyticFile_CS vars_shell program.sas
(Only for example)

libname in ";

data one;
set in.npcrfull2013_aid ;
css_flag=0;

*** Exclude non-reportable cases ***;
if ncprrptl=1;
if subm_st in ('PR','PI') then css_flag=9;

*** Exclude Invalid Age & unknown/other Sex ***;
if I230_AgeDx='999' then css_flag=2;
if I220_Sex='9' | I220_Sex='3' | I220_Sex='4' | I220_Sex='5' then css_flag=3;

*** ELIMINATE SINGLE FIELD EDIT ERRORS; ***
if E_CSF_AgeDx=1 | E_CSF_Site=1 | E_CSF_Race1=1 | E_CSF_Sex=1 then css_flag=4;

*** delete interfield edit errors ***;
if E_CIF_NAACCRIF13=1 | E_CIF_SEERIF15=1 | E_CIF_NAACCRIF47=1 | E_CIF_SEERIF17=1 then css_flag=5;

*** ICD-O-3 flags (new in 2004 submission mj); ***
if E_CIF_SEERMorph ge 1 | E_CIF_SEERIF39=1 | E_CIF_SEERIF38=1 | E_CIF_SEERIF15=1 then css_flag = 11;

*** delete records prior to reference year or 12-23 month old data ***;
if subm_st='AZ' | subm_st='CA' | subm_st='CO' | subm_st='GA' | subm_st='ID' | subm_st='IN' | subm_st='KY' | subm_st='MT' | subm_st='NH' | subm_st='NJ' | subm_st='NE' | subm_st='NV' | subm_st='PA' | subm_st='RI' | subm_st='WV' | subm_st='WA' then refyear='1995';
else if subm_st='AL' | subm_st='AK' | subm_st='DC' | subm_st='MS' | subm_st='MO' | subm_st='NY' | subm_st='OH' | subm_st='OR' | subm_st='SC' | subm_st='WY' then refyear='1996';
else if subm_st='AR' | subm_st='DE' | subm_st='ND' | subm_st='OK' then refyear='1997';
if dxyear<refyear then css_flag=7;

*** remove PR and PI; ***
if subm_st='PR' | subm_st='PI' then delete;

if I380_SeqNoCntrl in ('0 ','1 ') or I440_Grade = '0' then put "left in analysis, but invalid seqno or grade ";
I523_BehavICDO3= I522_HistTypeICDO3= I440_Grade= subm_st= dxyear= I380_SeqNoCntrl=;

hist2_s=substr(I420_Hist_9200_ICDO2,1,2);
hist3_s=substr(I522_HistTypeICDO3,1,2);

*** Re-define vital status to match SEER ***;
if I1760_VitalStatus='0' then I1760_VitalStatus='4';
sitenum=substr(I400_Site,2,3)+0;

*** combine insite and invasives for urinary bladder;
*** add codes for ICD-O-3 fields mj;
*** fix before create behanal variable;
if ('C670' <= I400_Site <= 'C679')and (I522_HistTypeICDO3 < '9590' or I522_HistTypeICDO3 > '9989') then do;
   I523_BehavICDO3 = 3;
   behanal = '3';
   if dxyear<2004 then I760_SS1977=1;
   if dxyear<2004 then I759_SS2000=1;
   urin3flag= 1;
end;

** recode astrocytomas to malignant **;
else if I522_HistTypeICDO3 in ('9421','9422') then behanal = '3';
else if I523_BehavICDO3=3 & ((8000<=I522_HistTypeICDO3<=8930 | 8932<=I522_HistTypeICDO3<=9132 | 9134<=I522_HistTypeICDO3<=9392 | 9394<=I522_HistTypeICDO3<=9537 | 9539<=I522_HistTypeICDO3<=9949 | 9951<=I522_HistTypeICDO3<=9959 | 9963<=I522_HistTypeICDO3<=9979 | I522_HistTypeICDO3=9988) | (I522_HistTypeICDO3=9133 & (.<sitenum<340 | sitenum>349))) then behanal='3';

** set to 4 any borderline to malignant histology - see Appendix 6 in ICD-O-3 **;
else if I523_BehavICDO3=1 & I522_HistTypeICDO3 in (8442,8451,8462,8472,8473) then behanal='5';
else if I523_BehavICDO3=2 then behanal = '2';
else behanal='9';

brthyear=input(substr(I240_DOB,5,4),$char4.);
brthmnth=input(substr(I240_DOB,1,2),$char2.);
fyyear=input(substr(I1750_DateLastContact,5,4),$char4.);
fumnth=input(substr(I1750_DateLastContact,1,2),$char2.);
primsite=input(substr(I400_Site,2,3),$char3.);
dxmnhn=substr(I1390_DateDx,1,2);dxyear=substr(I1390_DateDx,3,2);statenum=stfips(subm_st);

if statenum<10 then state=input('0'||trim(left(statenum)),$char2.);
else if statenum>=10 then state=input(trim(left(statenum)),Schar2);

if subm_st in ('KS','MN') then I90_CountyDx='000';

***fix CO counties;
if subm_st='CO' and dxyear<=2001 then do;
    if I90_CountyDx='001' then I90_CountyDx='911';
    else if I90_CountyDx='013' then I90_CountyDx='912';
    else if I90_CountyDx='059' then I90_CountyDx='913';
    else if I90_CountyDx='123' then I90_CountyDx='914';
end;

stcty=input(stfips||I90_CountyDx, $char5.);
state99=stfips;

*** 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;

*** if white, check ihslink ***;
if subm_st in ('AL','AZ','AR','CA','CO','DE','DC','FL','GA','ID','IL','IN','KS',
    'KY','LA','ME','MD','MA','MI','MN','MS','MO','MT','NE','NV','NH','NJ',
    'NY','NC','ND','OH','OK','OR','PA','RI','SC','SD','TN','TX','UT','VA',
    'WA','WV','WI','WY') then do;
    if racerec in ('1','5','9') & I192_IHS='1' then racerec='3';
end;

if I191_NHIA=0 then nhiaoth='0';
else if 1<=I191_NHIA<=8 then nhiaoth='1';

*** Create certification variable ***;
*** PW is Palau;
uscsstd = '1';
if dxyear = '2002' and subm_st in ('DC','MS','TN') then uscsstd = '0';
else if dxyear = '2001' and subm_st in ('MS','TN','VA') then uscsstd = '0';
else if dxyear = '2000' and subm_st in ('AR','MS','NC','SD','TN','VA') then uscsstd = '0';
else if dxyear = '1999' and subm_st in ('AR','MS','NC','SD','TN','VA') then uscsstd = '0';
else if dxyear = '1998' and subm_st in ('AR','GA','MD','MS','NC','NH','SD','TN','VA') then uscsstd = '0';

uscs9808 = '1';
if subm_st in ('AR','DC','GA','MD','MS','NC','NH','SD','TN','VA') then uscs9808 = '0';

uscs9908 = '1';
if subm_st in ('AR','DC','MS','NC','SD','TN','VA') then uscs9908 = '0';

uscs0408 = '1';

program='01';
retain cnt 0;
cnt+1;

if 1995<=fuyear<=2009 then nfuyear=fuyear;
else nfuyear=9999;

run;
Appendix II

NPCR_9915_AnalyticFile variables recode. Direct SEER*Stat extraction.

Behavior_recode_for_analysis

0 = "Benign"
1 = "Borderline malignancy"
2 = "In situ"
3 = "Malignant"
4 = "Only malignant in ICD-O-3"
5 = "No longer reportable in ICD-O-3"
6 = "Only malignant 2010+
15 = "Invalid Value(s)"

Site_recode_ICD_O_3_WHO_2008

1 = "Lip"
2 = "Tongue"
3 = "Salivary Gland"
4 = "Floor of Mouth"
5 = "Gum and Other Mouth"
6 = "Nasopharynx"
7 = "Tonsil"
8 = "Oropharynx"
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"
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"

ICCCsiterecextendedICDO3WHO200

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"
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"
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"
88 = "X(c.6) Malignant gonadal tumors of mixed forms"
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"

Race_recode_for_uses

1 = "White"
2 = "Black"
3 = "American Indian/Alaska Native"
4 = "Asian or Pacific Islander"
5 = "Other unspecified (1991+)")"
9 = "Unknown"
Sequence_number_central

0 = "One primary only"
1 = "1st of 2 or more primaries"
2 = "2nd of 2 or more primaries"
3 = "3rd of 3 or more primaries"
4 = "4th of 4 or more primaries"
5 = "5th of 5 or more primaries"
6 = "6th of 6 or more primaries"
7 = "7th of 7 or more primaries"
8 = "8th of 8 or more primaries"
9 = "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"
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"
56 = "56th of 56 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"
75 = "15th of 15 or more state registry-defined neoplasms"
80 = "20th 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"
99 = "Unknown seq num - federally required in situ or malig tumors"

SiterecwithKaposiandmesothelio

1 = "Lip"
2 = "Tongue"
3 = "Salivary Gland"
4 = "Floor of Mouth"
5 = "Gum and Other Mouth"
6 = "Nasopharynx"
7 = "Tonsil"
8 = "Oropharynx"
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"                     |
| 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)"

Stateraceethincl

0 = "Exclude state for race/ethnicity state-level analyses"
1 = "Include state for race/ethnicity state-level analyses"

Econ_status

0 = "Distressed"
1 = "At Risk"
2 = "Transitional"
3 = "Competitive"
4 = "Attainment"
9 = "Missing county data"
14 = "Blank(s)"

Merged_Radiation

1 = "had radiation"
2 = "did not have radiation"
3 = "patient or guardian refused radiation"
4 = "radiation recommended but unknown if received"
14 = "Blank(s)"
Merged_Summary_Stage

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"
8 = "Not applicable"
9 = "Unknown/unstaged/unspecified/DCO"
14 = "Blank(s)"