National Program of Cancer Registries Cancer Surveillance System (NPCR-CSS)

2018 Data Release Policy Diagnosis Years 1995–2017

Policy Revised July 2018

Cancer Surveillance Branch
Division of Cancer Prevention and Control
NCCDPHP, CDC
4770 Buford Hwy, N.E., Mailstop F-76
Atlanta, GA 30341-3717
E-mail: uscsdata@cdc.gov (specify "NPCR-CSS" in subject line)

TABLE OF CONTENTS

National Program of Cancer Registries
Cancer Surveillance System
2018 Data Release Policy
June 2018

I. INTRODUCTION

This document describes the format and content of data that the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) Cancer Surveillance System (CSS) releases or shares. This multi-year policy updates the July 2017 NPCR-CSS Data Release Policy. This policy applies to data submitted to the Centers for Disease Control and Prevention (CDC) for the 2018 NPCR-CSS data submission and for all future data submissions until a new policy is provided.

The NPCR-CSS Privacy Steward, as authorized by the Chief of the Cancer Surveillance Branch, clears all releases of state data, ensuring that the data are released according to the terms of the NPCR-CSS Data Release Policy.

It is possible that, in future years, data release practices or the content and format of released data may vary from those described in these guidelines. Such changes may occur as a result of improvements in the quality of the data, changes in information technology, and evolving data needs. However, if such variations occur, the data release practices will provide comparable protection (or more protection) for patient confidentiality to what is described in this policy. If it is anticipated that any data will be released with less protection (as determined by the NPCR-CSS Privacy Steward) for patient confidentiality than is described in this policy, NPCR central registries will be notified and have ample time to respond before the data are released. This policy is reviewed annually by the NPCR-CSS Privacy Steward and other appropriate CDC staff members to determine whether revisions are needed.

Summary of Changes

- Updated description of the USCS Data Visualization tool, page 4
- Information on IHS Data Visualization, page 6
- Information on access to NPCR and U.S. Cancer Statistics analytic datasets provided to the American Cancer Society, page 6
- Information on access to USCS Analytic Data through collaborative relationships and on-site access, page 8
- Information on the NPCR Prevalence and USCS Delay-Adjusted databases, page 9
- Updated threshold for cell suppression for the USCS Analytic Data, pages 12-13 and 20
- Description of procedure for external requests or access, page 14
- Updated list of primary sites provided to EPHTN, page 19
- Updated age groups and presentation of Summary Stage in USCS Data Visualization, page 20
- Updated data item list for USCS Analytic Data set, page 31
- Updated data item list for the NPCR/SEER USCS Public Use Dataset, page 40
- Updated data item list for the Restricted Access Dataset, page 50

II. OVERVIEW OF DATA

In 1992 Congress established NPCR by enacting the Cancer Registries Amendment Act, Public Law 102-515.[4] The law authorized CDC to provide funds and technical assistance to States and territories to improve or enhance existing cancer registries and to plan for and implement population-based central cancer registries where they did not exist. NPCR's purpose is to assure the availability of more complete local, state, regional, and national cancer incidence data for the planning and evaluation of cancer control interventions and for research. NPCR adopted reporting requirements and definitions consistent with the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results Program (SEER);[11,12] required the use of uniform data items, codes, and record layouts as defined by the consensus of members of the North American Association of Central Cancer Registries (NAACCR);[13] and established standards for data management and data completeness, timeliness, and quality similar to those recommended by NAACCR.[13,14] In 1994, the first 37 States received funding from CDC.[15] Currently, 47 States, the District of Columbia, Puerto Rico, Virgin Islands, and the U.S. Pacific Island Jurisdictions are funded by NPCR (appendix A).[16] NPCR-funded central registries collect data on patient demographics, primary tumor site, morphology, stage of disease at diagnosis, and first course of treatment. In addition, NPCR central registries conduct follow-up for vital status by linking with state and national death files or active case follow-up.

Invasive and *in situ* cancer case reports are submitted to CDC by population-based statewide central cancer registries in all 47 participating States, the District of Columbia, Puerto Rico, Virgin Islands, and the U.S. Pacific Island Jurisdictions. In each state or territory, state laws and regulations mandate the reporting of cancer cases by facilities and practitioners who diagnose or treat cancer to the state health department or its designee.[4] The central cancer registry receives case reports from facilities and practitioners throughout the state and processes them according to standard data management procedures.[14] Personal identifiers including the patient's name, Social Security number, and street address are removed from the NPCR-CSS submission prior to the encryption and electronic transmission of these case reports to a contractor acting on behalf of CDC. CDC and the contractor adhere to strict data security procedures when receiving, processing, and managing the data (appendix B). NPCR-CSS received formal approval (protocol #2594) from CDC's Institutional Review Board (IRB) in October 1999. The approval is updated annually. CDC has an Office for Human Research Protections (OHRP)-approved, federal-wide assurance of compliance with rules for the protection of human subjects in research (45 Code of Federal Regulations 46).

Central cancer registries and federal agencies routinely publish cancer incidence data 23 months after the close of each diagnosis year based on data that meet data quality standards.[16,17] However, other versions of the same data, based on the data file as it exists at different time periods, are usually available. For example, some central registries have preliminary data available as soon as 12 months after the close of each diagnosis year. After the publication of official statistics, central cancer registries (as well as CDC and NCI) continue to update and republish data with new information incorporated. When cancer incidence data are published, it is common practice to document either the data submission date (i.e., when the data were

submitted to CDC or NCI) or the date that the file was prepared. Changes in central cancer registry incidence data that occur more than 22 months after the close of a diagnosis year are likely to be small; however, delays in reporting are more likely to impact certain cancer sites and may be important for some research studies.[18]

CDC generates multiple data products using NPCR-only data and combined NPCR and NCI's Surveillance Epidemiology and End Results (SEER) data. The combined NPCR and SEER data are referred to as U.S. Cancer Statistics (USCS). USCS is the official federal cancer statistics, providing the most up-to-date information on the *entire* U.S. population.

III. DATA RELEASE ACTIVITIES

Starting with DP17-1701, participation in all CDC-created and hosted analytic datasets and web-based data query systems, as outlined in this policy, is a required strategy[1]. Therefore, the NPCR-CSS Dataset Participation Agreement is no longer provided.

A. Public Web-based Query Systems

For purposes of this policy, public web-based query systems are defined as datasets that are comprised of aggregated data (i.e., not individual case-specific data or microdata) that have been modified according to accepted procedures to block breaches of confidentiality and prevent disclosure of the patient's identity or confidential information and have a database behind a CDC firewall that is either case-specific microdata or pre-analyzed data tables.[2, 5–10] Users are able to access only aggregate counts and rates with all confidentiality protections built in. A combination of confidentiality protection measures is employed for each public web-based query system (see Table 1). These systems do not contain information that is identifiable or potentially identifiable according to currently accepted procedures for reducing disclosure risk.[2, 5–10] Before each system is finalized, the aggregate values are analyzed to determine whether there is a need for complementary cell suppression.[2, 5–10] If appropriate, the analysis includes consultation with a statistician with specific expertise in statistical disclosure limitation techniques. Following the analysis, complementary cell suppression is applied as needed.

There are no restrictions on access to public web-based query systems. A public release disclosure statement (see IV.C. Public Release Disclosure Statement) cautions users against inappropriate use of the data or inappropriate disclosure of information. Data are released as delimited ASCII files, a web-based query system, or possibly through other vehicles (see Table 1). As a convenience to NPCR central registries, state may request from CDC a copy of their complete state-specific analytic database that is used to populate each public web-based query system. The following public web-based query systems are currently being released:

1 DP17-1701, 2. CDC Project Description, a. Approach, iii. Strategies and Activities, Program 3: National Program of Cancer Registries (NPCR) – Component 1, Strategy 3 Cancer Data and Surveillance (Domain 1), Data Submission (page 19)

- USCS Data Visualizations Tool
- CDC WONDER – USCS incidence and NPCR survival data
- Federal partner's web-based query systems
    - NCI's State Cancer Profiles
    - Environmental Public Health Tracking Network (EPHTN)
    - Chronic Disease Indicators (CDI) website and data portal

All NPCR-CSS public web-based query systems consist of cancer incidence data selected from the NPCR/SEER analytic database. This is the same database that provides cancer incidence data for the annual release of USCS data products, including the Data Visualizations Tool, public use database and State Cancer Profiles. Data sources, case definitions, basic registry eligibility criteria in terms of required data quality, population denominator sources, methods for calculating incidence rates, and the rationale for specific cell suppression thresholds are as described in the USCS Data Visualizations Technical Notes, unless noted in separate documentation that accompanies the data.

Separate documentation may accompany each data product that describes its unique features (e.g., the data submission date, percentage of the U.S. population covered, diagnosis years and cancer sites included, variables included, data suppression rules, any special data quality criteria required for inclusion, and any unique statistical methods employed).

USCS Data Visualizations Tool

The USCS Data Visualizations Tool is a web-based application built with D3 Java Script libraries, written in Microsoft .NET, that outputs data in hypertext markup language (HTML) file containing the aggregate counts and rates for incidence, mortality, prevalence and survival estimates published annually, along with text documentation and data visualizations. The tool is available at www.cdc.gov/cancer/dataviz. It currently displays single year and 5-year aggregate counts, age-adjusted rates, and 95-percent confidence intervals by primary site, sex, race, and ethnicity at the county, state, regional, and national levels. Preliminary and delay-adjusted incidence rates and counts, as well as other newly identified indicators, may be published in the tool. The Data Visualizations tool has the database behind a CDC firewall with pre-tabulated data using SEER*Stat queries, which allows for the display of counts and rates. Users are able to access only aggregate counts and rates with all confidentiality protections built in. Downloadable ASCII files with the pre-tabulated data are available from the tool's website.

CDC WONDER – USCS Incidence, Mortality, Incidence/Mortality Ratios and NPCR Survival Data

The USCS dataset available on CDC WONDER displays the aggregate incidence and mortality counts, rates, and 95-percent confidence intervals, by primary site, sex, race, and ethnicity at the state, regional, Metropolitan Service Areas (MSA), and national levels. Cancer incidence/mortality ratios (by year, state, MSA, race, ethnicity, sex and cancer site) and 5-year relative survival data (NPCR-only data by race, sex, age group, and cancer site) are also available. The WONDER database is stored behind a CDC firewall with case-specific microdata. Users are only able to access only aggregate counts and rates with all confidentiality protections built in.

The WONDER tool allows users more flexibility in creating cross-tabulations than the Data Visualizations tool. While the same underlying USCS data is available in the two tools, more detailed breakdowns of counts and rates are available through WONDER. The additional values result from variable selections that are not currently available in the Data Visualizations tool (see Table 1) and include results for Metropolitan Service Areas that have met the population threshold of 50,000 or more and standard 5-year age groups that can be combined by the user.

Federal Partners' Web-based Systems

CDC shares aggregated data with federal partners for display in their web-based query systems. The data are generated specifically for the partners' needs and are shared via ASCII files. Unless otherwise noted below, the data generally consists of aggregate cancer incidence counts, crude rates, and age-adjusted rates for selected primary sites, age groups, and counties in the United States (see Table 1 for more details).

Future versions may contain more detail about cancer at the county level. Beginning in 2008, CDC began routinely publishing county data averaged over 5 years.

Age-adjusted rates only

State Cancer Profiles is a web-based query tool that public health professional can use to prioritize cancer control efforts at the county-, state-, and national-level. Data released to NCI SEER for the State Cancer Profiles data product includes age-adjusted incidence and mortality rates only.

Age-adjusted and crude rates

Data released to the U.S. Department of Health and Human Services, Office of Women's Health (OWH) includes crude and age-adjusted rates. The data is available through their online tool, Health Information Gateway.

Environmental Public Health Tracking Network

USCS data are provided to the CDC's National Center for Environmental Health's Environmental Health Tracking Network (EPHTN). The EPHTN portal displays single-year and 5-year aggregate incidence counts, age-adjusted rates, and 95-percent confidence intervals for selected primary sites and age groups at the state and county level (see Table 1). Single-year can be viewed at the state-level; data by 5-year average and 5-year summed are available at the county-level. The EPHTN web-based query system runs using a database behind a CDC firewall with case-specific microdata, which allows for the calculation of locally-weighted smoothed rates or unsmoothed rates, or both:

- **EPHTN Unsmoothed Rates**
  Data published are similar to that on State Cancer Profiles. It includes cancer data from all 50 states.
- **EPHTN Smoothed Rates**
  Smoothing is the process of averaging a measure for an area based on information about that area and areas around it. Please note that the main purpose of smoothing is to clarify spatial patterns and to improve the stability of rates, not to prevent disclosure of private

information. Back-calculation of case counts from smoothed rates is sometimes possible when the method of smoothing is made known and (non-sensitive) denominator data are available from other sources.

Through EPHTN, users are able to access only aggregate counts and rates with all confidentiality protections built in.

- **EPHTN National Portal to State Portal**
  CDC's Environmental Health Tracking Branch (EHTB) has grantees in several NPCR-funded states that are responsible for the state-level public portals. In collaboration with EHTB, upon request, CDC-NPCR provides the state-level EPHTN dataset to the EHTB state counterpart.

Indian Health Services (IHS)

CDC continues to use the IHS linkage results for analyses of cancer incidence among American Indians/Alaska Natives (AI/AN). In addition to improving cancer incidence rates presented in USCS Data Visualization, an analytic database is maintained by a CDC Division of Cancer Prevention and Control employee assigned to IHS. Access to this database is limited to approved CDC staff. The data is used to respond to data requests for AI/AN cancer incidence rates from tribal epidemiology centers and tribal organizations contingent upon permission from the state registries that comprise the IHS areas of interest. By December 2018, 5-year aggregate incidence counts, age-adjusted rates, and 95-percent confidence intervals for selected primary sites at the IHS will be displayed in USCS Data Visualizations tool (see Table 1) Inclusion in this dataset also allows IHS to ***provide the state with the date of death obtained through NDI-IHS linkage and/or the date the linkage occurred by diagnosis year, for registries that complete an NDI supplemental confidentiality agreement for application Y9-0033***.

B. Data Release to Federal and Trusted Partners

American Cancer Society (ACS)

CDC shares NPCR and USCS data with ACS in order to promote collaborations on cancer surveillance and epidemiological research efforts. ACS's Surveillance and Health Services Research (SHSR) Program analyzes and disseminates cancer statistics and identifies gaps and opportunities for cancer prevention, early detection and treatment. The SHSR annually publishes the statistical report, *Facts and Figures*, and peer-reviewed journal articles that are used by public health experts, clinicians, and scientists.

In 2018, a Memorandum Of Understanding was implemented with the American Cancer Society, and ACS staff members must sign a Data Use Agreement form and complete annual Assurance of Confidentiality training before s/he is given access to the data. CDC provides ACS staff access to the following databases with record level data through SEER*Stat software: USCS delay-adjusted database, NPCR survival database, NPCR prevalence database, and selected variables from the NPCR and SEER Quality Control database. The Quality Control database shared with SEER is restricted to 24-month data, excludes postal code and census tract variables, and excludes "day" fields for date of birth and date of death.

Central Brain Tumor Registry of the United States (CBTRUS)

CBTRUS annually publishes the print and Web versions of the statistical report, *Primary Brain Tumors in the United States Statistical Report Supplement*; a previous version of the report is available at: http://www.cbtrus.org/reports/reports.html. The report includes age-adjusted rates and corresponding 95-percent confidence intervals on brain and other central nervous system tumors and is presented by state, histology, major histology grouping, primary site, behavior, gender, race, ethnicity, and age at diagnosis. CDC provides individual, record-level data to CBTRUS for the publication of this report; Appendix C lists the variables included in this dataset. Only states meeting the USCS publication criteria are included in the dataset.

In addition, CBTRUS uses these data to respond to inquiries that are more specific than those that are provided by the report. For these inquiries, no individual record level data is released; only aggregated data with the corresponding confidence intervals (if applicable) and appropriate suppression criteria are provided to data inquirers. Attribution to the NPCR is provided. CBTRUS signs data use agreements before data are released for their report and future inquiries. For questions, contact CBTRUS staff at cbtrus@aol.com.

International Association of Cancer Registries (IACR)

The International Association of Cancer Registries (IACR) produce the *Cancer Incidence in Five Continents* (CI5) and the *International Incidence of Childhood Cancer* (IICC). The *CI5* series of monographs, published every five years, has become the reference source of data on the international incidence of cancer. The most recent version was published in 2017. The CI5 databases provide access to detailed information on the incidence of cancer recorded by cancer registries (regional or national) worldwide in two formats (CI5 and CI5plus) and the IICC provides access to detailed information on the incidence of pediatric cancers:

- ***CI5***
  Presents the basic data published in the CI5 volumes.
- ***CI5plus***
  Contains annual incidence for selected cancer registries published in CI5 for the longest possible period.
- ***IICC***
  Presents basic pediatric data.

When IACR requests data, the formal Call for Data Submission giving information on the evaluation procedure, likely layout of how data will be presented, and questionnaire on registry operations will be available from the IACR website. CDC-NPCR will provide additional information regarding the CI5 Call for Data as it becomes available. There are two components of the CI5 Call for Data: 1) the questionnaire and introductory text and 2) data submission.

Data submitted for CI5 may also be used for the IICC publication making a separate data submission unnecessary. This IACR product does require a separate questionnaire and introductory text to be completed by the states.

States are responsible for completing the on-line questionnaires and providing an introductory text, indicating if the CI5 data and introductory text are also used for the IICC product. CDC-

NPCR will submit aggregated NPCR data for central cancer registries meeting USCS publication criteria.

CONCORD

CONCORD is the global program for world-wide surveillance of cancer survival and is led by the London School of Hygiene & Tropical Medicine and supported by the Union for International Cancer Control (UICC). CONCORD monitors progress towards the overarching goal of the UICC World Cancer Declaration made in 2013: "major reductions in premature deaths from cancer, and improvements in quality of life and cancer survival".

A call for participation in the CONCORD studies is periodically issued and extends examination of world-wide cancer survival trends for certain cancer sites: i.e., stomach, colon, rectum, liver, lung, breast, cervix, ovary, prostate, esophagus, pancreas, and melanoma of skin in adults, as well as leukemias, lymphomas, and brain tumors in adults and children (0-14 years). The protocol and dataset specifications are posted to NPCR-CSS Document Server, CONCORD tab as they become available.

CDC-NPCR submits NPCR data for central cancer registries meeting USCS publication criteria for survival analyses (meet USCS data quality criteria and have conducted active patient follow-up or linked records with the National Death Index).

Agency for Healthcare Research and Quality (AHRQ)

Health and Human Service's Agency for Healthcare Research and Quality (AHRQ) is the lead federal agency charged with improving the safety and quality of America's health care system. It develops and disseminates knowledge, tools, and data to improve health care systems and help Americans, health care professionals, and policy makers make informed health decisions. NPCR-CSS data are shared with AHRQ for reports on national healthcare quality and disparities.

C. Analytic datasets

USCS Analytic Data

CDC creates USCS Analytic Datasets each year that include data from central cancer registries meeting USCS publication criteria and diagnosis year coverage. CDC, NCI staff members, and contractors perform analyses of USCS data as needed using these internal analytic databases created using the USCS data – that is created from combined NPCR and SEER Program data.

The datasets are made available via SEER*Stat software to federal employees, fellows, and contractors in the Division of Cancer Prevention and Control and NCI's Surveillance, Epidemiology, and End Results Program (SEER) after signing a *NPCR Analytic Data Use Agreement* (Appendix D) and *CDC Nondisclosure Agreement* (appendix E) and completing annual Assurances of Confidentiality training. The dataset is also available to approved partnering organizations and state central cancer registries after a Memorandum of Understanding and Data Use Agreements are signed (see Appendix H and Appendix I).

In specially-established collaborative relationships, researchers external to CDC, NCI, and ACS may be provided access to the USCS analytic datasets. In these relationships, CDC staff must be

included in the analytic project as a co-author, Data Use Agreements must be signed, and Assurance of Confidentiality training must be completed before access will be provided. Additionally, access will only be allowed on-site at CDC's Cancer Surveillance Branch offices. See the section "External Data Requests".

Cancer surveillance and epidemiological analyses include assessment of the completeness, timeliness, and quality of cancer incidence data and analyses of the cancer burden and survival as needed for meeting national cancer control objectives. Such analyses of state and national data are conducted routinely by federal agencies, including CDC and SEER, for programmatic or statistical purposes, as needed, to achieve the agencies' mandates.

There are four internal analytic datasets routinely analyzed by CDC and SEER staff members:

NPCR/SEER USCS Incidence Analytic Dataset
CDC, NCI staff members, and contractors conduct cancer surveillance and epidemiological research that result in publications, data briefs, and presentations. Examples of research include descriptive analyses by racial and ethnic populations for specific cancers, descriptions of cancer incidence trends, and descriptive analyses of the quality of the data. Appendix F lists the variables available in this dataset.

NPCR Internal Survival Dataset
Cancer survival data is critical for evaluating the progress and impact of early detection/screening programs and/or comprehensive cancer control plans as well as interventions from other sources. CDC's NPCR-CSS calculates and publishes survival estimates on this population at the national, state, and regional levels. Focusing on the entire NPCR-CSS dataset supports analyses of survival estimates for rare cancers that cannot be addressed otherwise and provides data for publication on the USCS website as official statistics for the U.S. Appendix G lists the variables available in this dataset.

NPCR Internal Prevalence Dataset
This database provides 5-year limited duration prevalence estimates for NPCR registries who meet USCS publication criteria for all years included in the database and that have completed National Death Index linkages or active patient follow-up for all years included in the database. The list of variables available in this dataset are in Appendix O.

NPCR/SEER USCS Delay-Adjusted Dataset
Case-reporting delay may result in an underestimate of true incidence. Researchers can adjust for this delay using composite delay factors, thus producing more precise cancer incidence trends. The composite delay factors used in this database were developed by SEER and are used by NPCR, SEER, and NAACCR. The delay-adjustment factors account for cancer site, registry, age, race, ethnicity, and diagnosis year, and are used to estimate delay-adjusted counts and rates. The list of variables available in this dataset are in Appendix P.

In compliance with the 308(d) Assurance of Confidentiality, CDC and SEER employees and contractors and partner organizations conducting these analyses are required to handle the information in accordance with principles outlined in the CDC Staff Manual on Confidentiality

and to follow the specific procedures documented in the NPCR-CSS Confidentiality/Security Statement (appendices B, H, and I).

In addition, CDC, SEER, and partner organization staff members are required to acknowledge state cancer registries whenever NPCR-CSS data are presented, released, or published by CDC by making available the following (or similar) statement:

> *These data were provided by central cancer registries participating in the National Program of Cancer Registries (NPCR) and submitted to CDC in [Month, Year], and/or the Surveillance, Epidemiology and End Results (SEER) program and submitted to NCI in [Month, Year]. The dataset includes data for diagnosis years 1998-xxxx (excluding SEER-Metro Registry data).*

NPCR/SEER USCS Incidence Public-Use Research Dataset

For purposes of this policy, the NPCR/SEER USCS Incidence Public-Use Research Dataset is defined as the version of the full NPCR/SEER USCS microdata (i.e., individual case-specific data) that have been modified as needed to minimize the potential for disclosure of confidential information. It consists of a subset of data items published in USCS. This dataset does not contain personal identifiers such as a patient's name, street address, or Social Security number as this information is not transmitted by central cancer registries to CDC as part of their annual data submission. Certain data items, such as date of birth, and reporting-source (death certificate only and autopsy) cases have also been removed from this research dataset to minimize the potential identification of individuals with the occurrence of rare cancer in a person of certain age or racial or ethnic group or living in a specific county. The list of the variables included in this dataset is in Appendix J.

The dataset, previously only available to NPCR Registry Staff, is now available publicly through SEER*Stat software. Researchers are given access to the data after signing an *NPCR and SEER – U.S. Cancer Statistics Research Data Use Agreement* (Appendix K) and SEER Research Data Use Agreement (https://seer.cancer.gov/data/access.html). A Public Release Disclosure Statement cautions users against inappropriate use of the data or inappropriate disclosure of information. Cell suppression of <16 cases is automatic and the SEER*Stat case listing function is disabled as additional data protection measures. This dataset allows the authorized counts, crude rates, age-adjusted incidence rates, and 95-percent confidence intervals to be generated by the authorized user to meet their specific needs.

Restricted-Access Research Dataset (RDC)

For purposes of this policy, the restricted-access dataset is defined as the version of the full NPCR/SEER USCS analytic dataset, either aggregated data or microdata (i.e., individual case-specific data) that has been modified as needed to minimize (but may not remove entirely) the potential for disclosure of confidential information.

CDC uses the National Center for Health Statistics Research Data Center (NCHS RDC) as a mechanism for researchers outside of the Division of Cancer Prevention and Control (DCPC) to request and gain access to NPCR data for research purposes. The data is available through the NCHS RDC only after the standard data quality reviews that occur as part of the preparation for

USCS and State Cancer Profiles. The restricted-access dataset is released to researchers through the NCHS RDC after CDC authenticates the requestor's identity and research intent through an extensive proposal review process and after the researcher completes the NCHS RDC confidentiality and security requirements. The requestor must also comply with the confidentiality procedures at and data sharing agreements with the NCHS RDC.

The NCHS RDC has developed and maintains detailed data sharing agreements and procedures for user authentication and for logging and monitoring of data releases. Proposed project proposals are reviewed by staff at central cancer registries, through the NPCR Central Cancer Registry Council, and by CDC, which includes NPCR and NCHS RDC staff. User documentation includes a data dictionary for every diagnosis year available at the NCHS RDC.

The use of the NCHS RDC to manage data access provides the highest level of data security and protection of confidentiality that is available for data analysis. Using the NCHS RDC allows CDC to comply with the Assurance of Confidentiality [308(d)] that was obtained for the NPCR-CSS data. The NCHS RDC is also covered by a separate Assurance of Confidentiality [308(d)].

For further information regarding the NCHS RDC, refer to Appendix L of this policy.

The restricted-access dataset does not contain personal identifiers such as a patient's name, street address, or Social Security number as this information is not transmitted by central cancer registries to CDC as part of their annual data submission. However, the dataset may contain information that is potentially identifiable especially when linked with other datasets, such as the occurrence of a rare cancer in a person of a certain age or racial or ethnic group or living in a specific county. The data is made available to researchers through a SAS dataset. The RDC staff creates a SAS dataset specific to each project. Researchers must include a data dictionary in their proposal and only the requested variables are included in the SAS file.

D. Data Release Under Controlled Conditions

CDC-wide policy stipulates that a CDC program may consider release of data that cannot be released as either a public web-based system, a research dataset, or restricted-access dataset under certain controlled conditions.[1] These controlled conditions may include a CDC-controlled data center such as the data center established at National Center for Health Statistics (NCHS) (http://www.cdc.gov/nchs/r&d/rdc.htm), on-site at CDC's Cancer Surveillance Branch offices, or through special licensing. Except as described above, NPCR-CSS data will not otherwise be released under these controlled conditions while the current policy is in place. Release of data under controlled conditions will be considered as part of discussions with partners, and a determination will be made as to whether such releases of data will be considered for NPCR-CSS data.

E. Emergency and Provisional Data Releases

It is not anticipated that CDC will need to release NPCR-CSS data before the files have been modified as needed to protect confidentiality as described in this policy. This is prohibited by the 308(d) Assurance of Confidentiality (appendices B, H, and I).

Provisional data and draft data tables may be shared with CDC employees and contractors, NPCR central registries, and other partners in order to facilitate quality reviews of the data.

When appropriate, individuals who participate in such reviews sign a *NPCR Analytic Data Use Agreement* and a *CDC Nondisclosure Agreement* (when applicable) before accessing the data or tables.

IV. PROTECTION OF DATA

A. Assurance of Confidentiality

All data collected and maintained by NPCR-CSS must be managed, presented, published, and released with strict attention to confidentiality and security, consistent with the general principles and guidelines established by CDC for confidential case data[1–3] and specific restrictions imposed on NPCR-CSS data (appendices B, H, and I).[4] Special care must be given to cancer incidence data that are not directly identifiable because geographic and small cell data may be indirectly identifying when combined with detailed information in case reports, laboratory reports, medical records, or linkage with other data files.[5–10]

NPCR-CSS has approval for protection under section 308(d) of the Public Health Services (PHS) Act (42 U.S.C. 242m(d)) (appendices B, H, and I). The 308(d) confidentiality assurance protects identifiable and potentially identifiable information from being used for any purpose other than the purpose for which it was collected (unless the person or establishment from which it was obtained has consented to such use). This assurance protects against disclosures under a court order and provides protections that the Privacy Act of 1974 (5 U.S.C. 552a) does not. For example, the Privacy Act of 1974 protects individual participants, but the 308(d) confidentiality assurance also protects institutions. Confidentiality protection granted by CDC promises participants and institutions that their data will be shared only with those individuals and institutions listed in the project's consent form or in its specified policies.

B. Suppression of Rates and Counts

When the numbers of cases or deaths used to compute rates are small, those rates tend to have poor reliability. Another important reason for using a threshold value for suppressing cells is to protect the confidentiality of patients whose data are included in a report by reducing or eliminating the risk of disclosing their identity.

Therefore, to discourage misinterpretation or misuse of rates or counts that are unstable because case or death counts are small, annual incidence and death rates and counts in publicly available datasets and web-based query systems are suppressed if the case or death counts are below 16. A count of fewer than about 16 results in a standard error of the rate that is approximately 25% or more as large as the rate itself. Similarly, a case count below 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 assumed to be similar to the observed population age distribution. For aggregated time periods, counts and rates are suppressed for less than 16 cases. However, average annual rates and counts may not be suppressed if the total case count for the time period exceeds 16.

The cell suppression threshold value of 16, which was selected to reduce misuse and misinterpretation of unstable rates and counts, is more than sufficient to protect patient confidentiality.

Asian/Pacific Islander and American Indian/Alaskan Native data are presented only for the nation, and states with at least 50,000 population, because of concerns regarding possible misclassification of race data and the relatively small sizes of these populations in the United States.

Per the Data Use Agreements, researchers using internal analytic files are required to suppress case counts less than 6 in publications and presentations. Researchers are advised to use caution when presenting or interpreting results based on less than 16 cases.

Complementary cell suppression and suppression of certain race and ethnicity combinations are required as additional measures to assure confidentiality and stability.

C. Public Release Disclosure Statement

The following (or similar) public release disclosure statement is prominently displayed for users of all NPCR-CSS public web-based query systems, research datasets, and restricted-access datasets:

> *Data Use Restrictions: Read Carefully Before Using*
> ***By using these data, you signify your agreement to comply with the following statutorily based requirements.*** *The National Program of Cancer Registries (NPCR), Centers for Disease Control and Prevention (CDC), has obtained an assurance of confidentiality pursuant to Section 308(d) of the Public Health Service Act, 42 U.S.C. 242m(d). This assurance provides that identifiable or potentially identifiable data collected by the NPCR may be used only for the purpose for which they were obtained unless the person or establishment from which they were obtained has consented to such use. Any effort to determine the identity of any reported cases, or to use the information for any purpose other than statistical reporting and analysis, is a violation of the assurance. Therefore users will:*
>
> - *Use the data for statistical reporting and analysis only.*
> - *Make no attempt to learn the identity of any person or establishment included in these data.*
> - *Make no disclosure or other use of the identity of any person or establishment discovered inadvertently, and advise the Associate Director for Science, Office of Science Policy and Technology Transfer, CDC, Mailstop D-50, 1600 Clifton Road, N.E., Atlanta, Georgia, 30333, Phone: 404-639-7240) (or NCI's SEER Program if SEER data) and the relevant state or metropolitan area cancer registry, of any such discovery.*

D. Freedom of Information Act (FOIA) Data Requests

The Freedom of Information Act (FOIA) (http://www.cdc.gov/od/foia/) generally provides that, upon written request from any person, a federal agency (i.e., CDC) must release any agency record unless that record falls (in whole or part) within one of nine exemptions. FOIA applies to federal agencies only and covers only records in the possession and control of those agencies at the time of the FOIA request (except in certain instances involving grantee-held data). Because state-based data become a federal record in CDC's possession, such records are subject to

disclosure in response to a FOIA request. The FOIA exemptions that may be available to protect some aspects of state data from public disclosures in response to a FOIA request are:

- Exemption 3, which specifically exempts information from disclosure by statute (in this instance, pursuant to an Assurance of Confidentiality under Section 308(d) of the Public Health Service Act), and
- Exemption 6, which exempts from disclosure personnel and medical files and similar files, which would constitute an unwarranted invasion of personal privacy.

In general, non-FOIA requests to CDC from the public, media, and other government agencies for local cancer incidence data are referred to the state health department for a reply. There are three reasons for this: (1) the state health departments can release cancer incidence data in accordance with locally established policies and procedures and consistent with provisions of the Cancer Registries Amendment Act (Public Health Service Act, (42 USC 280e-280e-4), as amended);[4] (2) the relative infrequency of data submission to federal agencies assures that the state health department or its designated central cancer registry will have the most complete, accurate, and up-to-date information; and (3) the central registry may be able to provide more detailed data that can better meet the needs of the requestor. When the request is for data regarding cancer incidence involving more than one state, CDC will refer the requestor to published reports or to NPCR-CSS datasets that are released in accordance with practices described in this document, if relevant.

E. CDC External Data Requests

Individuals, agencies, or organizations outside CDC may request data not available from a public web-based query system or research dataset. When the requests do not identify a state, CDC staff members or contractors tabulate the data for the inquirer. For requests that identify a state, CDC staff members may seek States' permission regarding use. See Appendix N for additional details.

Researchers may submit data query or study proposal requests for the NPCR/SEER USCS Incidence Analytic Dataset to CDC. These requests must include:

- Names of individuals who will need access to the data
- Purpose and public health significance of the investigation
- Research question(s)
- Variables required beyond those in the freely-available research data
- Subset of cases needed (specifically cancer type, data years, registries)
- Planned use of data (e.g., manuscript, poster, presentation)

After CDC authenticates the requestor's identity and research intent, and verifies that confidentiality is maintained, a CDC analyst will process the data query and provide results to the researcher. The requestor must comply with all confidentiality and data suppression procedures outlined in the NPCR-CSS Assurance of Confidentiality [308(d)].

In circumstances where the researcher requires access to the USCS Analytic Datasets:

- CDC staff must be included in the analytic project as a co-author
- Data Use Agreements must be signed
- Assurance of Confidentiality training must be completed

- Access is only allowed on-site at CDC's Cancer Surveillance Branch offices.

V. REFERENCES

1. Centers for Disease Control and Prevention. *CDC/ATSDR Policy on Releasing and Sharing Data*. Atlanta: Centers for Disease Control and Prevention; 2003. Available at http://www.cdc.gov/maso/Policy/ReleasingData.pdf.
2. Centers for Disease Control and Prevention. *CDC/ATSDR/CSTE Data Release Guidelines for Re-Release of State Data*. Atlanta: Centers for Disease Control and Prevention; 2003 (available upon request).
3. Centers for Disease Control and Prevention. *CDC Staff Manual on Confidentiality*. Atlanta: Centers for Disease Control and Prevention; 1984 and National Center for Health Statistics. *NCHS Staff Manual on Confidentiality*. Hyattsville, MD: National Center for Health Statistics; 1999.
4. Cancer Registries Amendment Act, Public Law 102-515, Stat. 3312 (October 22, 1992). Available at http://www.cdc.gov/cancer/npcr/npcrpdfs/publaw.pdf.
5. American Statistical Association. Data Access and Personal Privacy: Appropriate Methods of Disclosure Control. Alexandria, VA: American Statistical Association; 2008. Available at http://www.amstat.org/news/statementondataaccess.cfm.
6. Doyle P, Lane JI, Theeuwes JM, Zayatz LM (eds). *Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies*. Amsterdam: Elsevier Science BV; 2001.
7. Federal Committee on Statistical Methodology. *Checklist on Disclosure Potential of Proposed Data Releases*. Available at http://fcsm.sites.usa.gov/committees/cdac/cdac-checklist/.
8. Federal Committee on Statistical Methodology. *Report on Statistical Disclosure Limitation Methodology*. (*Statistical Working Paper* 22). Washington, DC: Office of Management and Budget; 1994. Available at http://www.fcsm.gov/reports/#fcsm.
9. McLaughlin C. Confidentiality protection in publicly released central cancer registry data. *Journal of Registry Management* 2002; 29(3):84–88.
10. Stoto M. *Statistical Issues in Interactive Web-Based Public Health Data Dissemination Systems*. Draft report prepared for the National Association of Public Health Statistics and Information Systems, Rand Corporation; September 2002.
11. Surveillance, Epidemiology, and End Results Program. *The SEER Program Code Manual*. 3rd ed. Bethesda, MD: National Cancer Institute,1998.
12. Percy C, Van Holten V, Muir C (eds). *International Classification of Diseases for Oncology, 2nd edition.* Geneva, Switzerland: World Health Organization; 1990.
13. Hultstrom D, editor. *Standards for Cancer Registries, vol. II: Data Standards and Data Dictionary*, version 9.1, 6th ed. Springfield, IL: North American Association of Central Cancer Registries; 2001.

14. North American Association of Central Cancer Registries. *Standards for Cancer Registries, vol. III: Standards for Completeness, Quality, Analysis, and Management of Data*. Springfield, IL: North American Association of Central Cancer Registries; 2000.

15. Hutton MD, Simpson, LD, Miller DS, Weir HK, McDavid K, Hall HI, Progress toward nationwide cancer surveillance: an evaluation of the National Program of Cancer Registries, 1994–1999. *Journal of Registry Management* 2001;28(3):113–120.

16. U.S. Cancer Statistics Working Group. *U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report, Technical Notes.* Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2015. Available at: www.cdc.gov/uscs.

17. NAACCR Method to Estimate Completeness. A Data Analysis Tool for Calculations. Available at: http://www.naaccr.org/Research/DataAnalysisTools.aspx.

18. Clegg LX, Fueur EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. *Journal of the National Cancer Institute* 2002; 94(20):1537–45.

TABLE 1 –Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Overview						
	Public Web-Based Query Systems				Analytic datasets	
	USCS Data Visualizations Tool	**USCS WONDER[2]**	**USCS Data for Partners[3]**	**EPHTN**	**USCS Public-Use Research Database**	**USCS Restricted-Access Dataset**
Format	Database of aggregate counts and rates, with text documentation	Database of aggregate counts and rates, with text documentation. The database behind the CDC firewall is case-specific microdata.	Database of aggregate counts and rates, with text documentation	Database of aggregate counts and rates, with text documentation. The database behind the CDC firewall is case-specific microdata.	Customized, analytic database. The database behind the SEER*Stat firewall is case-specific microdata with enforced cell suppression and case listing disabled.	Customized, analytic database available through proposal process
Mode of Access	Web-based query system with downloadable ASCII files, MS Excel files, and SAS datasets	Web-based query system	Flat ASCII file, web-based query system, and separate brief text documentation	Web-based query system	SEER*Stat client-server mode only after receipt of signed Data Use Agreement	On-site at CDC or through CDC staff assistance
Web Address or Contact Information	USCS Web site www.cdc.gov/cancer/dataviz	CDC WONDER http://wonder.cdc.gov	Request from uscsdata@cdc.gov (specify "USCS County" in subject line)	National Environmental Public Health Tracking http://www.cdc.gov/nceh/tracking	https://www.cdc.gov/cancer/ public-use	Application process available at www.cdc.gov/rdc
Contains Potentially Identifiable Information	No		No	No	No	Yes
Registry Eligibility Criteria for Data Completeness and Quality	USCS publication criteria		USCS publication criteria; data meet criteria for unknown county	USCS publication criteria; data meet criteria for unknown county	USCS publication criteria	USCS publication criteria; data meet criteria for unknown county
When Available	Updated 2018		Updated 2018	Updated 2018	Updated 2018	Updated 2018

2 This data file is also shared with OWH.
3 This data file is shared with CDI and AHRQ.

TABLE 1 – Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Cases Included						
	Public Web-Based Query Systems				Analytic datasets	
	USCS Data Visualizations Tool	**USCS WONDER**	**NPCR/SEER USCS County**	**EPHTN**	**USCS Public-Use Research Database**	**USCS Restricted-Access Dataset**
States/ Territories	NPCR/SEER States meeting eligibility criteria		NPCR/SEER States meeting eligibility criteria	NPCR States meeting eligibility criteria	NPCR/SEER States meeting eligibility criteria	NPCR States meeting eligibility criteria
Diagnosis Years	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2014; 2015; 2011-2015; 2016; 2017 preliminary results	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2014; 2015; 2016	2011-2016	Individual years 2001 through 2015 for state level; 2001-2005, 2002-2006, 2003-2007, 2004–2008; 2005-2009; 2006-2010; 2007-2011; 2008-2012; 2009-2013; 2010-2014; 2011-2015; 2012-2016 county level	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2014; 2010-2014; 2011-2015; 2012-2016	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2014; 2015; 2016
Cancer Sites	All reportable invasive cancers; *in situ* female breast, and benign and borderline primary intracranial and central nervous system tumors (diagnosis year 2004)		All reportable cancer sites combined; female breast; *in situ* female breast; cervix uteri; colon and rectum; lung and bronchus; melanoma; bladder; prostate; oral cavity and pharynx; brain and other nervous system; thyroid; kidney and renal pelvis; stomach; ovary; corpus and uterus, NOS; leukemias; non-Hodgkin lymphoma; liver and intrahepatic bile duct; pancreas, esophagus; and childhood cancers	Female breast; lung and bronchus; bladder; brain & other nervous system; thyroid; leukemias (all types; Acute myeloid leukemia; Chronic lymphocytic leukemia); non-Hodgkin lymphoma; all childhood cancers (state level only); childhood leukemias (state level only); childhood CNS & miscellaneous intercranial & intraspinal neoplasms (state level only); mesothelioma (state level only); kidney & renal pelvis; prostate; melanoma of skin; liver & intrahepatic bile duct; pancreas; oral/pharynx; esophagus, larynx; testicular	All reportable invasive cancers; *in situ* female breast, and benign and borderline primary intracranial and central nervous system tumors (diagnosis year 2004)	All reportable invasive and *in situ* cancers and benign and borderline primary intracranial and central nervous system tumors (diagnosis year 2004)

TABLE 1 – Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Variables Included						
	Public Web-Based Query Systems				Analytic datasets	
	USCS Data Visualizations Tool	**USCS WONDER**	**USCS Data for Partners**	**EPHTN**	**USCS Public-Use Research Database**	**USCS Restricted-Access Dataset**
Geographic Levels	All areas combined; U.S. census region and division; NPCR/SEER state, territory, county; SEER metropolitan area, IHS regions (AI/AN data only)	All areas combined; NPCR and SEER state or territory; MSA for cities of ≥500,000 (additional levels may be added)	NPCR and SEER state or territory; county	NPCR state; county	All areas combined; U.S. census region and division; NPCR and SEER state or territory	NPCR and SEER state or territory; county for approved requests only
Race/Ethnicity	All races combined; white; black; Asian/Pacific Islander (API); American Indian/Alaska Native (AI/AN); Hispanic; white Hispanic; white non-Hispanic; black Hispanic; black non-Hispanic		All races combined; white; black; AI/AN; API (with appropriate 50,000 population suppression); Hispanic; white/black Hispanic/non-Hispanic	All races combined; white; black; AI/AN; API (with appropriate 50,000 population suppression); Hispanic	All races combined; white; black; Asian/Pacific Islander (API) ; American Indian/Alaska Native (AI/AN); Hispanic; white Hispanic; white non-Hispanic; black Hispanic; black non-Hispanic	All races reported; Hispanic; white Hispanic; white non-Hispanic; black Hispanic; black non-Hispanic
Age Groups	All ages combined and standard 5-year age groups for adults and <15,<20, and 5-year age groups for childhood cancers	All ages combined and standard 5-year age groups that can be combined by the user	Childhood cancers: <15 and <20; all other cancers: <50, 50–64, 65+	Childhood cancers: <15 and <20 Breast cancer: <50, 50+	All ages combined, standard 5-year age groups	Standard 5-year age groups and individual ages (Month and day of birth not provided for confidentiality reasons. If the age at diagnosis >99, then grouped into one category. Year of birth is also grouped.)
Summary Stage	Yes		No	No	Yes	Yes
Histology	International Classification of Childhood Cancers, Third Revision (all geographic areas combined), Mesothelioma (national and level), Kaposi Sarcoma (national and level), Consensus Conf on Cancer Registration of Brain, and CNS Tumors (all geographic areas combined)		No	No	Same as USCS	Yes

TABLE 1 – Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Confidentiality Protection/Disclosure Limitation Measures Employed						
	Public Web-Based Query Systems				Analytic datasets	
	USCS Data Visualizations Tool	USCS WONDER	NPCR/SEER USCS County	EPHTN	USCS Public-Use Research Database	USCS Restricted-Access Dataset
Direct or Record-Level Identifiers?	No		No	No	No	Yes, but not in output which will be reviewed by CDC staff for confidentiality
Aggregation	Yes		Yes	Yes	No	No
Limited Number of Variables	Yes		Yes	Yes	Yes	Yes
Grouping/Collapsing of Variables or Response Codes; e.g., race and age recode	Yes		No	Yes	No	Yes
(1) Average Annual Counts Rounded to the Nearest Whole Number (2) Average Annual Rates (3) Annual Averages Are Based on At Least 5 Years of Data	Yes for county presentation		Yes	Yes	No	No
Cell Suppression	Yes Counts and rates: count of <16		Yes Counts and rates: 5 year total count of <16	Yes Counts and unsmoothed rates: count of <16 Smoothed rates: RSE >25%	Yes Counts and rates: count of <16 enforced Case listing disabled	Yes (output reviewed by CDC analyst to ensure counts of <6 are suppressed)
Complementary Cell Suppression	As needed		As needed	As needed	As needed	As needed
Public Release Disclosure Statement	Yes		Yes	Yes	Yes	Yes
Data Sharing Agreement and/or IRB Approval	No		No	No	Yes	Yes
User Authentication	No		No	No	No	Yes
Logging and Monitoring	Limited		Limited	Limited	Yes, monitoring databases used, session type and date only	Yes

APPENDIX A – State and Metro Area Cancer Registries

State, Metropolitan Area, and Territory Cancer Registries by Federal Funding Source, and First Diagnosis Year***** **for Which Cancer Cases Were Reportable to CDC's NPCR or NCI's SEER Program**

State, Metropolitan Area, or Territory	First Diagnosis Year for Which Cancer Cases Were Reportable to NPCR or SEER*	Federal Funding Source
Alabama	1996	NPCR
Alaska	1996	NPCR
Arizona	1995	NPCR
Arkansas	1996	NPCR
California	1995/2000	NPCR/SEER
Los Angeles	1992	SEER
San Francisco-Oakland	1973	SEER
San Jose-Monterey	1992	SEER
Colorado	1995	NPCR
Connecticut	1973	SEER
Delaware	1997	NPCR
District of Columbia	1996	NPCR
Florida	1995	NPCR
Georgia	1995/2010	NPCR/SEER
Atlanta	1975	SEER
Hawaii	1973	SEER
Idaho	1995	NPCR
Illinois	1995	NPCR
Indiana	1995	NPCR
Iowa	1973	SEER
Kansas	1995	NPCR
Kentucky	1995/2000	NPCR/SEER
Louisiana	1995/2000	NPCR/SEER
Maine	1995	NPCR
Maryland	1996	NPCR
Massachusetts	1995	NPCR
Michigan	1995	NPCR
Detroit	1973	SEER
Minnesota	1995	NPCR
Mississippi	1996	NPCR
Missouri	1996	NPCR
Montana	1995	NPCR
Nebraska	1995	NPCR
Nevada	1995	NPCR
New Hampshire	1995	NPCR

APPENDIX A – State and Metro Area Cancer Registries

State, Metropolitan Area, and Territory Cancer Registries by Federal Funding Source, and First Diagnosis Year*** for Which Cancer Cases Were Reportable to CDC's NPCR or NCI's SEER Program**

State, Metropolitan Area, or Territory	First Diagnosis Year for Which Cancer Cases Were Reportable to NPCR or SEER*	Federal Funding Source
New Jersey	1995/2000	NPCR/SEER
New Mexico	1973	SEER
New York	1996	NPCR
North Carolina	1995	NPCR
North Dakota	1997	NPCR
Ohio	1996	NPCR
Oklahoma	1997	NPCR
Oregon	1996	NPCR
Pennsylvania	1995	NPCR
Puerto Rico	1998	NPCR
Rhode Island	1995	NPCR
South Carolina	1996	NPCR
South Dakota	2000	NPCR
Tennessee	1999	NPCR
Texas	1995	NPCR
United States Pacific Island Jurisdictions	2007	NPCR
Utah	1973/2016	SEER/NPCR
Vermont	1996	NPCR
Virginia	1996	NPCR
Virgin Islands	2016	NPCR
Washington	1995	NPCR
Seattle-Puget Sound	1974	SEER
West Virginia	1995	NPCR
Wisconsin	1995	NPCR
Wyoming	1996	NPCR

* Diagnosis year is the year during which a reported cancer case was first diagnosed.

CDC = Centers for Disease Control and Prevention
NCI = National Cancer Institute
NPCR = National Program of Cancer Registries
SEER = Surveillance, Epidemiology, and End Results Program

APPENDIX B – NPCR-CSS Overview of Data Security

The NPCR-CSS project data reside on a dedicated server maintained by the NPCR-CSS contractor. To ensure the security and confidentiality of project data, the following provisions have been incorporated into the NPCR-CSS Security Plan in accordance with the requirements of the Assurance of Confidentiality.

The NPCR-CSS server is housed in a secure facility with a guard on duty 24 hours a day. Only authorized staff is allowed to access the facility. Support people are escorted by an authorized staff member if needed. The server resides on its own local area network (LAN) behind the NPCR-CSS contractor's firewall. NPCR-CSS contractor project staff access the server via VPN from their primary office location. Elevator and stairwell access is controlled by card key 24 hours. During business hours, an attendant is always present at the reception desk to guide visitors.

- Access to the NPCR-CSS server is limited to authorized NPCR-CSS contractor project staff. It is password-protected on its own security domain. No one, including NPCR-CSS contractor non-project staff, is allowed access to the NPCR-CSS data.
- All NPCR-CSS contractor project staff must sign a confidentiality agreement before passwords and keys are assigned. All staff must pass background checks appropriate to their responsibilities for a public trust position.
- NPCR-CSS data that are submitted electronically are encrypted during transmission from the States. They arrive on a document server behind the NPCR-CSS contractor's firewall. Each state has its own directory location so that no state has access to another state's data. The data are moved automatically from the document server to the NPCR-CSS server.
- Receipt and processing logs are maintained to document data receipt, file processing, and report production. All reports and electronic storage media containing NPCR-CSS data are stored under lock and key when not in use and will be destroyed once they are no longer needed.
- A comprehensive security plan has been developed by the NPCR-CSS contractor's security team. The security team consists of the Project Director, Project Manager, Systems Lead and Security Officer, Database Administrator and LAN/WAN Security Steward. All project staff receive annual security awareness training covering security procedures. The ICF International project security team oversees operations to prevent unauthorized disclosure of the NPCR-CSS data.
- Periodic (currently quarterly, but no less than once per year) reviews and updates of the NPCR-CSS contractor's security processes will be conducted to adjust for rapid changes in computer technology and to incorporate advances in security approaches. The Security Plan will be amended as needed to maintain the continued security and confidentiality of NPCR-CSS data.

APPENDIX C – Data Items for CBTRUS

The dataset for CBTRUS includes individual case-specific data from the NPCR-CSS dataset. The data items to be included are listed below.

**Diagnosis Years 1995-2016 invasive cases only, 2004-2065 invasive, benign, and borderline cases*

Item Name	NAACCR Data Item Number	Comments
Patient ID (unique)	20	
NAACCR Record Version	50	
State of Residence at Diagnosis	80	
County at Diagnosis	90	*Results presented as 5-year average annual rates as the smallest time period with <16 cell and complementary cell suppression required*
Rural/Urban Continuum/Beale Code 1993	3300	
Rural/Urban Continuum/Beale Code 2003	3310	
NPCR Race Recode	Derived based on [160], [161], and [192]	*Same as race for USCS*
NHIAv2 Derived Hispanic Origin	191	
NAPIIA	193	
Sex	220	
Age at Diagnosis	230	*Single year up to age 84; 85+ grouped into one category*
Sequence Number—Central	380	
Date of Diagnosis (*YEAR portion only*)	390	*Day and month of diagnosis not requested*
Date of Diagnosis (full date)	390	Full date
Primary Site	400	
Laterality	410	
Grade	440	
Diagnostic Confirmation	490	
Type of Reporting Source	500	
Histologic Type (ICD-O-3)	522	
Behavior (ICD-O-3)	523	
SEER Summary Stage 1977	760	
SEER Summary Stage 2000	759	

APPENDIX C – Data Items for CBTRUS

Item Name	NAACCR Data Item Number	Comments
Derived Summary Stage 2000	3020	
NPCR Cancer Stage		*Based on 759 and 3020*
RX Summ--Surgery Primary Site	1290	*2003-2015 diagnosis years*
RX Summ—Radiation	1360	*2003-2015 diagnosis years*
Rad–Regional RX Modality	1570	*2003-2015 diagnosis years*
Merged Radiation		*Based on 1360 and 1570* *1 = had radiation* *2 = did not have radiation* *3 = patient or guardian refused radiation* *4 = radiation recommended but unknown if received* *Applied only for selection below:* **8000**<=I522_HistTypeICDO3<=**9049** | **9056**<=I522_HistTypeICDO3<=**9139** | **9141**<=I522_HistTypeICDO3<=**9589**
EDITS overrides	1990–2074	
CS Site-Specific Factor 1	2880	*WHO Grade*

APPENDIX D – NPCR/SEER USCS Analytic Data Use Agreement

U.S Cancer Statistics Analytic Data

Submitted [Month,Year] (diagnosis years 1998-xxxx)

To protect the confidentiality of the individuals represented within the National Program of Cancer Registries – Cancer Surveillance System (NPCR-CSS) data, the Centers for Disease Control and Prevention (CDC) has obtained an Assurance of Confidentiality under Section 308(d) of the Public Health Service Act (42 U.S.C. 242m(d)), which provides that these data can only be used for the purpose for which they were obtained.

When using U.S. Cancer Statistics analytic data for research purposes, it is absolutely necessary to ensure, to the extent possible, that use of the data will be limited to research or public health purposes. In accordance with applicable federal law, there must be no attempt to determine the identity of individuals represented by reported cases, or to use the information for any purpose other than for health statistical reporting and analysis.

CDC's Division of Cancer Prevention and Control (DCPC) takes every possible measure to ensure that the identity of data subjects cannot be determined. All direct identifiers, as well as characteristics that might lead to identification of individuals, are omitted from the dataset. Certain demographic and clinical information has been included for research purposes; thus, all results must be presented or published in a manner that ensures that no individual can be identified. In addition, there must be no attempt to identify individuals from any computer file or to link with a computer file containing patient identifiers.

Data users must agree to the following provisions prior to receiving access to U.S. Cancer Statistics Incidence, U.S. Cancer Statistics Delay Adjusted, NPCR Prevalence and/or NPCR Survival Analytic Data. ***Please initial after each statement to indicate agreement.***

As the recipient of the U.S. Cancer Statistics Incidence (diagnosis years {year}-{year}), U.S. Cancer Statistics Delay Adjusted (diagnosis years {year}-{year}), NPCR Prevalence (diagnosis years {year}-{year}), and/or NPCR Survival* Analytic *Data (diagnosis years {year}-{year}):

- I will adhere to the requirements of the Data Use Agreement and understand that my access to the data will be revoked if these requirements are violated. **Initials: _____**
- I understand that the U.S. Cancer Statistics Incidence, U.S. Cancer Statistics Delay Adjusted, NPCR Prevalence and NPCR Survival Analytic Data belong to the states and territories. The states' and territories' agreement to use of the data is obtained through the activities outlined in the general NPCR-CSS Data Release Policy and by specific requests to the states and territories through the CSB management team.
  Initials: _____
- I will not use or permit others to use the datasets in any way other than for statistical reporting and analysis. **Initials: _____**

APPENDIX D – NPCR/SEER USCS Analytic Data Use Agreement

- I will not release or permit others to release the datasets or any part of them to any person except with DCPC's written approval. **Initials:** _____
- I will not attempt to link or permit others to link the datasets with individually identifiable records from any other dataset without DCPC's approval. **Initials:** _____
- I will not attempt to use the datasets or permit others to use them to learn the identity of any person or establishment included in any dataset. **Initials:** _____
- I will protect the data file(s) I receive with a password and/or encryption. In addition, any temporary or permanent analysis files, such as those produced with analytic software, will be protected in the same manner(s). **Initials:** _____
- I will take the following actions if the identity of any person or establishment is discovered inadvertently:
    - Make no use of this knowledge.
    - Notify DCPC's Cancer Surveillance Branch (CSB) Chief.
    - As requested by DCPC, safeguard or destroy the information that identifies an individual or establishment
    - Inform no one else of the discovered identity. **Initials:** _____
- I understand that calculating rates or other statistics based on small numbers can raise statistical issues concerning stability and confidentiality. I will use appropriate caution when presenting and interpreting results based on less than 16 cases. **Initials:** _____
- I agree that all oral or written reports will contain only aggregate data and no report of the data containing cells with less than 6 cases will be released. **Initials:** _____
- I will use complementary cell suppression to ensure that no data on an identifiable case can be derived through subtraction or other calculation from the combination of tables in all oral and written presentations. **Initials:** _____
- I have reviewed and am familiar with the Assurance of Confidentiality Training documentation posted on the Internal Data Users Group's intranet site. **Initials:** _____
- I have added my project to the NPCR Internal Analysis SharePoint table and, if applicable, I will notify and obtain permission from the Internal Data Users Group to analyze state- and county-level data. **Initials:** _____
- I will acknowledge central cancer registries whenever data are presented, released, or published by including the following (or similar) statement:

    These data were provided by central cancer registries participating in the National Program of Cancer Registries (NPCR) and submitted to CDC in November {year}, and/or the Surveillance, Epidemiology and End Results (SEER) program and submitted to NCI in November {year}. The U.S. Cancer Statistics Incidence Analytic dataset includes diagnosis years {year}–{year} (excluding SEER-Metro Registry data); U.S. Cancer Statistics Delay Adjusted Analytic dataset includes diagnosis years {year}–{year} (excluding SEER-Metro Registry data), NPCR Prevalence Analytic dataset includes diagnosis years {year}–{year} and the NPCR Survival Analytic dataset includes diagnosis years {year}–{year}.
    Initials: _____

- As appropriate, I will cite the data:

  *National Program of Cancer Registries SEER*Stat Database: {Database file name} – {year}-{year}. United States Department of Health and Human Services, Centers for Disease Control and Prevention. Released {date}, based on the November {year} submission.*

  Initials: _____

- I understand that if I require technical assistance in analyzing or interpreting the data and when such assistance goes beyond providing non-manipulated data, IDUG members reserves the right to request to be considered as a research collaborator or co-author in any resulting publications or presentations. **Initials: _____**
- I will provide a courtesy copy of draft papers or abstracts to the NPCR Internal Data Users Group at npcridug@cdc.gov as they are entered into Documentum for clearance. **Initials: _____**
- I am familiar with the use of **SEER*Stat** in analyzing data or will complete the needed training. **Initials: _____**

My signature below indicates that I agree to comply with all the above stated provisions.

_____

Signature Date

Name:_____

Title_____

Branch_____

Telephone_____ E-mail:_____

Please return completed form to the NPCR Internal Data Users Group at npcridug@cdc.gov.

APPENDIX E – CDC Non-Disclosure Agreement

The success of CDC's operations depends upon the voluntary cooperation of States, of establishments, and of individuals who provide the information required by CDC programs under an assurance that such information will be kept confidential and be used only for epidemiological or statistical purposes.

When confidentiality is authorized, CDC operates under the restrictions of Section 308(d) of the Public Health Service Act which provides in summary that no information obtained in the course of its activities may be used for any purpose other than the purpose for which it was supplied, and that such information may not be published or released in a manner in which the establishment or person supplying the information or described in it is identifiable unless such establishment or person has consented.

"I am aware that unauthorized disclosure of confidential information is punishable under Title 18, Section 1905 of the U.S. Code, which reads:

'Whoever, being an officer or employee of the United States or of any department or agency thereof, publishes, divulges, discloses, or makes known in any manner or to any extent not authorized by law any information coming to him in the course of his employment or official duties or by reason of any examination or investigation made by, or return, report or record made to or filed with, such department or agency or officer or employee thereof, which information concerns or relates to the trade secrets, processes, operations, style of work, or apparatus, or to the identity, confidential statistical data, amount or source of any income, profits, losses, or expenditures of any person, firm, partnership, corporation, or association; or permits any income return or copy thereof or any book containing any abstract or particulars thereof to be seen or examined by any person except as provided by law; shall be fined not more than $1,000, or imprisoned not more than one year, or both; and shall be removed from office or employment.'

"I understand that unauthorized disclosure of confidential information is also punishable under the Privacy Act of 1974, Subsection 552a (i) (1), which reads:

'Any officer or employee of an agency, who by virtue of his employment or official position, has possession of, or access to, agency records which contain individually identifiable information the disclosure of which is prohibited by this section or by rules or regulations established thereunder, and who knowing that disclosure of the specific material is so prohibited, willfully discloses the material in any manner to any person or agency not entitled to receive it, shall be guilty of a misdemeanor and fined not more than $5,000.'

''My signature below indicates that I have read, understood, and agreed to comply with the above statements.''

Typed/Printed Name Signature Date

Center/Institute/Office

APPENDIX F – Data Items for NPCR/SEER USCS Incidence Analytic Dataset

SEER*Stat Category	SEER*Stat Variable Name
Age at Diagnosis	Age recode with <1 year olds
Race, Sex, Year Dx, Registry, County	Sex
	Year of diagnosis
	Addr at DX – state
	*County at DX
	*State-county
	USCS standard
	USCS9815
	Race recode for USCS
	Program
	*Econ status
	*Region/Division
	Region
	USCS9915
	USCS0615
	USCS1115
	Origin recode NHIA (Hispanic, Non-Hisp)
Site and Morphology	Behavior Recode for Analysis
	Primary Site – labeled
	*Primary Site
	Histologic Type ICD-O-3
	*Behavior Code ICD-O-3
	Grade
	Diagnostic confirmation
	ICD-O-3 Hist/behavior, labeled
	*ICD-O-3 Hist/behavior, malig, labeled
	Site recode ICD-O-3/WHO 2008
	ICCC site recode ICD-O-3/WHO 2008
	ICCC site rec extended ICD-O-3/WHO 2008
	AYA site recode/WHO 2008
	Lymphoma subtype recode/WHO 2008
	Behavior recode for analysis derived/WHO2008
Stage – LRD [Summary and Historic]	*Derived SS2000
	*SEER Summary Stage 2000
	*SEER Summary Stage 1977
	Merged Summary Stage 2000
Therapy	*RX summ – surg prim site
	*RX summ – chemo
	*Merged radiation
Extent of Disease – CS	*CS extension

APPENDIX F – Data Items for NPCR/SEER USCS Analytic Dataset

	*CS lymph nodes
	*CS mets at dx
	*CS site-specific factor 1
	*CS site-specific factor 2
	*CS site-specific factor 3
	*CS site-specific factor 15
	Laterality
Multiple Primary Fields	Sequence number - central
Race and Age (case data only)	Age at Diagnosis
	*Race 1
	*IHS Link
Geographic Locations	*Ruralurban continuum 2013
	*Census Tract Poverty Indicator
	*Ruralurban continuum 2013 calc
Dates	Year of Birth
	Month of diagnosis
Other	Type of Reporting Source
Merged System-Supplied	Alcohol-related cancers
	HPV-related cancers
	Obesity-related cancers
	Physical inactivity-related cancers
	Tobacco-related cancers
	State race eth suppress

* Variable is only available in the internal incidence database; it is not available in the NPCR Public Use Database

APPENDIX G – Data Items for NPCR Internal Survival Dataset

SEER*Stat Category	SEER*Stat Variable Name
Age at Diagnosis	Age recode with single ages and 85+
Race, Sex, Year Dx, Registry, County	Sex
	Year of diagnosis
	Addr at DX – state
	County at DX
	State-county
	NPCR project flag
	Economic status 2015
	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)
	Race recode (White, Black, Other)
	Year of diagnosis
Site and Morphology	Primary Site – labeled
	Histologic Type ICD-O-3
	Behavior Code ICD-O-3
	Grade
	Diagnostic confirmation
	ICD-O-3-Hist/behavior, labeled
	ICD-O-3-Hist/behavior, malig, labeled
	Site recode ICD-O-3/WHO 2008
	ICCC site recode ICD-O-3/WHO 2008
	Behavior recode for analysis derived/WHO2008
Stage – LRD [Summary and Historic]	Derived SS2000
	SEER Summary Stage 2000
	Merged Summary Stage 2000
Extent of Disease – CS	CS Site-Specific Factor 1
	CS Site-Specific Factor 2
	CS Site-Specific Factor 15
	Laterality
Cause of Death (COD) and Follow-up	Survival months – presumed alive
	Survival months flag – presumed alive
	Cause of death (ICD-10)
	ICD revision number
	Vital status
	Follow-up source central
	COD exclusion flag
	Original vital status
	Vital status recode (study cutoff used)
	Cause of death recode
	COD recode with Kaposi and mesothelioma

APPENDIX G – Data Items for NPCR Internal Survival Dataset

SEER*Stat Category	SEER*Stat Variable Name
Multiple Primary Fields	Sequence number - central
Race and Age (case data only)	Age at Diagnosis
	Race 1
	NHIA derived Hispanic origin
Dates	Presumed alive year of last contact recode
	Presumed alive month of last contact recode
	Presumed alive day of last contact recode
	Year of birth
	Month of diagnosis
	Day of diagnosis
	Original day of last contact
	Original month of last contact
	Original year of last contact
	Original year of diagnosis
	Original day of diagnosis
	Original month of diagnosis
Other	Type of Reporting Source
User-Specified	EDPMDE LinkVar

APPENDIX H – NPCR-CSS 308(d) Assurance of Confidentiality Statement

A surveillance system of population-based cancer incidence data received from cooperative agreement holders for the National Program of Cancer Registries is being conducted by the National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) of the Centers for Disease Control and Prevention (CDC), an agency of the U.S. Department of Health and Human Services, and ICF, a contractor of CDC. The information to be received by CDC is a subset of a standard set of data items that the state central cancer registry routinely receives from hospitals, pathology labs, clinics, and private physicians on all cancer patients diagnosed in the state. This information includes patient demographics and cancer diagnosis and treatment data. Each year, CDC requests cumulative data from central cancer registries. The variables reported to CDC may vary from year to year. The cancer registries maintain these data permanently in longitudinal databases that are used for public health surveillance, program planning and evaluation, and research. CDC updates its longitudinal database each year with data received from the States. These data are used by CDC scientists for routine cancer surveillance, program planning and evaluation, and to provide data for research. NCCDPHP, recognizing the sensitivity of the data being furnished by the states, has applied for and obtained an Assurance of Confidentiality to provide a greater level of protection for the data while at CDC and at the contractor site.

Information received by CDC or its contractors as part of this surveillance system that could lead to direct or indirect identification of cancer patients is collected and maintained at CDC under Section 306 of the Public Health Service (PHS) Act (42 U.S.C. 242k) with an assurance that it will be held in strict confidence in accordance with Section 308(d) of the PHS Act (42 U.S.C. 242m). It is used only for purposes stated in this assurance and are not otherwise disclosed or released, even following the death of cancer patients in this surveillance system.

Information collected by CDC is used without personal identifiers for publication in statistical and analytic summaries and for release in restricted release datasets for research. Information that could lead to direct or indirect identification of cancer patients is not made available to any group or individual. In particular, such information is not disclosed to: insurance companies; any party involved in civil, criminal, or administrative litigation; agencies of federal, state, or local government; or any other member of the public.

Collected information that could lead to direct or indirect identification of cancer patients is kept confidential and—with the exception of CDC employees, their contractors, and qualified researchers—no one is allowed to see or have access to the information. CDC employees and contractors are required to handle the information in accordance with principles outlined in the CDC Staff Manual on Confidentiality and to follow the specific procedures documented in the Confidentiality Security Statement for this project. Qualified researchers are required to sign the NCHS RDC data sharing agreements and abide by the NCHS RDC confidentiality procedures. Organizations (e.g., the North American Association of Central Cancer Registries, American Cancer Society, and National Cancer Institute) are required to sign a detailed data release agreement to have access to restricted release data.

APPENDIX I – NPCR-CSS 308(d) Assurance of Confidentiality FAQ

Background

The Centers for Disease Control and Prevention (CDC) is responsible for public health surveillance in the United States. CDC collects, compiles, and publishes a large volume of personal, medical, epidemiologic, and statistical data. The success of CDC's operations depends, in part, on the agency's ability to protect the confidentiality of these data. While it is a matter of principle for CDC to guard sensitive information and federal statutes such as the Privacy Act of 1974 provide a degree of protection for personally identifiable data, Section 308(d) of the Public Health Service Act (42 U.S.C. 242m(d)) enables CDC to provide the highest level of confidentiality protection for sensitive and mission-significant research and surveillance data.

CDC received a formal delegation of authority from the National Center for Health Statistics (NCHS) (formally a separate agency) to grant 308(d) confidentiality protection in 1983. Section 308(d) of the Public Health Service Act ensures the confidentiality of data collected under Sections 304 and 306 of the Public Health Service Act. These special legislative authorities were the provisions under which NCHS collects and safeguards most of its survey data, along with the mortality data within the National Death Index. CDC was required to establish a stringent application process and continues to use the authority sparingly. The agency has granted confidentiality assurances to projects deemed significant to CDC's mission, such as surveillance of hospital infections, AIDS and HIV infections, pregnancy-related mortality, and congenital defects. Fewer than 65 projects have received 308(d) protection since CDC received this authority, and currently there are approximately 20-5 active projects with 308(d) confidentiality assurances. As a testament to the importance of this project to the mission of CDC, the National Program of Cancer Registries (NPCR) has been afforded this special data protection.

What is stated in Public Health Service Act, Section 308(d)?

The first clause of Section 308(d) states that CDC must explain the purpose for collecting data to persons or agencies supplying information, and it guarantees that CDC will be limited to those specified uses unless an additional consent is obtained. Moreover, the information obtained may be used only by CDC staff or CDC's contractors in the pursuit of such stated purposes. The second clause states that CDC may never release identifiable information without the advance, explicit approval of the person or establishment supplying the information or by the person or establishment described in the information.

What process did NPCR undertake to obtain 308(d) confidentiality protection?

NPCR staff worked with the CDC Office of General Counsel and the CDC Confidentiality and Privacy Officer to prepare the application for the NPCR Cancer Surveillance System (CSS) project. The application contained the following four components:

- A Justification Statement summarizing the NPCR-CSS project's programmatic purpose, the type of data to be collected, and the uses to be made of the information. This statement also included an assurance that a) the requested data would not be furnished without the guarantee of a confidentiality assurance, b) confidentiality assurance is important to protect the individuals described in the data and to reassure the institutions

submitting data, c) the information cannot reliably be obtained from other sources, d) the information is essential to the project's success, e) granting the confidentiality assurance would not prohibit CDC from fulfilling its responsibilities, and f) the advantages of assuring confidentiality outweigh the disadvantages.

- An Assurance of Confidentiality Statement delineating anticipated data uses and those with whom identifiable data would be shared, along with general advisements regarding the confidentiality protection.
- A Confidentiality Security Statement detailing the stringent safeguarding measures in place to ensure that the promise of confidentiality would not be jeopardized by practices of staff handling the data.
- An Institutional Review Board (IRB) Review Status Statement verifying NPCR-CSS's exemption from CDC IRB approval. (The Human Subjects Administrator at the National Center for Chronic Disease Prevention and Health Promotion determined that NPCR-CSS activities are routine surveillance and not research on human subjects. Therefore, protocol review by CDC IRB was deemed unnecessary.)

The application was submitted to the CDC Confidentiality Officer for review and modification, prepared for presentation to the CDC Confidentiality Review Group (CRG), and in May 2000 NPCR received 308(d) confidentiality protection approval for NPCR-CSS data, including authorization for retroactive confidentiality protection beginning with diagnosis year 1995. NPCR must file for continuation every 5 years to maintain the assurance. In 2006, 2010, and 2015, NPCR filed and received approval for continuation.

What makes 308(d) confidentiality assurance the best protection for NPCR-CSS data?
The 308(d) confidentiality assurance is the only confidentiality protection that covers routine surveillance activities, such as those conducted by NPCR-CSS. The assurance specifies that data protected by 308(d) may be used only for statistical or epidemiological purposes and not released further in identifiable form without consent. Another exclusive advantage of 308(d) is that it also protects indirectly identifiable data. Operationally, this means that NPCR may never release a directly identifiable variable (e.g., Social Security number) or any combination of variables that could be used to indirectly identify an individual. Finally, 308(d) provides protection for information on both living and deceased individuals.

Are there any disadvantages to individuals or institutions protected by the 308(d) confidentiality assurances?
A 308(d) confidentiality assurance does not pose a disadvantage for individuals or institutions submitting data to CDC. In fact, 308(d) provides an added benefit because it prevents CDC from freely releasing data to researchers and any other persons or entities that could request access to the data. With the confidentiality assurance protecting NPCR-CSS data, NPCR staff members are prohibited from sharing data except for the purposes stated at the time of data collection, unless consent from those who provided the assurance is obtained.

Does NPCR's 308(d) confidentiality assurance protect the data from subpoena and Freedom of Information Act (FOIA) requests?

APPENDIX I – NPCR-CSS 308(d) Assurance of Confidentiality FAQ

The 308(d) assurance is the strongest protection against compulsory legal disclosure that CDC can offer. Although CDC receives FOIA requests, the FOIA (b)(6) exemption enables CDC to withhold sensitive, individually identified data that would constitute a "clearly unwarranted invasion of personal privacy." It is CDC's firm position that all projects covered by a 308(d) confidentiality assurance, including NPCR-CSS, meet this exemption.

Has a case involving 308(d) been tested in court?
Yes. CDC's ability to protect data submitted to the agency was upheld in court. The case involved a National Institute for Occupational Safety and Health project collecting death certificate information, which is widely accepted as the least sensitive data protected by 308(d). The court's ruling in favor of the non-release of these data establishes an effective precedent for restricting access to more sensitive data, such as that collected by a cancer registry.

How long are confidential data submitted to NPCR-CSS protected?
NPCR-CSS data are covered by the 308(d) confidentiality assurance forever. Individual records in the NPCR-CSS surveillance system are protected even following the death of the cancer patients.

Will NPCR release CSS data to persons or agencies outside of CDC?
An assurance of confidentiality protects NPCR-CSS data held at CDC and by its contractor. The 308(d) confidentiality protection does not go with the data whether released publically or through restricted means, and any data released to qualified researchers by CDC are subject to the limits of any coverage afforded by the requesting agency. However, it is important to note that NPCR's confidentiality assurance prohibits the release of any data that are directly or indirectly identifiable. Therefore, CDC would not release highly sensitive NPCR-CSS data. Restricted access data that are released to external researchers are done so in accordance with the NCHS RDC proposal process and confidentiality procedures, prohibiting attempts to identify subjects within the record system. Under the 308(d), NPCR is permitted to release NPCR-CSS data to qualified researchers and organizations, such as the North American Association of Central Cancer Registries (NAACCR), American Cancer Society (ACS), and National Cancer Institute (NCI). This is so because these entities were specifically mentioned in the NPCR-CSS confidentiality assurance as anticipated recipients of identifiable data. Prior to the restricted release of NPCR-CSS data to qualified organizations, a detailed data use agreement must be signed by the requesting party (attachment I). Information that could lead to the identification of cancer patients, through direct or indirect methods, cannot be made available to any other group or individual. In particular, NPCR cannot disclose information to insurance companies; any party involved in civil, criminal, or administrative litigation; agencies of federal, state, or local government; or any other member of the public.

Are there penalties for violating the confidentiality assurance?
NPCR employees and NPCR-CSS contractor staff working on the NPCR-CSS project may be subject to fine, imprisonment, and termination of employment for unauthorized disclosure of confidential information. To assure that all NPCR employees are aware of their responsibilities to maintain and protect NPCR-CSS records and the penalties for failing to comply, CDC

employees must read and sign a data use agreement. Contract employees with access to NPCR-CSS data are required to sign a confidentiality agreement.

APPENDIX J – Data Items for NPCR/SEER USCS Incidence Public Use Research Dataset

The research use NPCR/SEER USCS Incidence Public Use dataset contains individual case-specific data from the USCS dataset with enforced <16 cell suppression and case listing disabled.

SEER*Stat Category	SEER*Stat Variable Name
Age at Diagnosis	Age recode with <1 year olds
Race, Sex, Year Dx, Registry, County	Sex
	Year of diagnosis
	Addr at DX – state
	USCS standard
	Race recode for USCS
	Program
	Region
	USCS0115
	USCS0615
	USCS1115
	Origin recode NHIA (Hispanic, Non-Hisp)
Site and Morphology	Primary site – labeled
	Histologic type ICD-O-3
	Grade
	Diagnostic confirmation
	ICD-O-3 hist/behavior, labeled
	Site recode ICD-O-3/WHO 2008
	ICCC site recode ICD-O-3/WHO 2008
	ICCC site rec extended ICD-O-3/WHO 2008
	AYA site recode/WHO 2008
	Lymphoma subtype recode/WHO 2008
	Behavior recode for analysis derived/WHO2008
Stage – LRD [Summary and Historic]	Merged summary stage 2000
Extent of Disease – CS	Laterality
Multiple Primary Fields	Sequence number – central
Race and Age (case data only)	NHIA derived Hisp origin
Dates	Year of birth
	Month of diagnosis
Other	Type of reporting source
Merged System-Supplied	State race eth suppress
	Alcohol-related cancers
	HPV-related cancers
	Obesity-related cancers
	Physical inactivity-related cancers
	Tobacco-related cancers

APPENDIX K – NPCR Research Data Use Agreement

National Program of Cancer Registries (NPCR) and Surveillance, Epidemiology, and End Results (SEER) Incidence – U.S. Cancer Statistics
Public Use Research Database Data Use Agreement
For data submitted November, {year}

The Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI) make NPCR and SEER data available to the public and researchers through various data release activities. The NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Databases are an unrestricted subset of data submitted to CDC and SEER and made available only through the National Cancer Institute's SEER*Stat statistical software.

CDC has obtained an assurance of confidentiality for NPCR pursuant to Section 308(d) of the Public Health Service Act, 42 U.S.C. 242m(d). Any effort to determine the identity of any reported cases, or to use the information for any purpose other than statistical reporting and analysis, is a violation of the assurance. All direct identifiers, as well as characteristics that might easily lead to identification of individuals, are omitted from the NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Databases. Certain demographic information has been included for research purposes; thus, all SEER*Stat results must be presented or published in a manner that ensures that no individual can be identified. In addition, there must be no attempt to identify individuals from any computer file or to link with a computer file containing patient identifiers.

Data users must agree to the following provisions prior to receiving access to the NPCR and SEER Incidence – U.S. Cancer Statistics {year}–{year} and {year}– {year} Public Use Research Databases. ***Please initial after each statement to indicate agreement.***

As the recipient of access to NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Databases:

- I will adhere to the requirements of the Data Use Agreement and understand that my access to the data will be revoked if these requirements are violated. **Initials:** _____
- I understand that all NPCR data are owned by the states and territories. The states and territories have established agreements with CDC regarding the use and dissemination of the data. **Initials:** _____
- I will not use or permit others to use the analytic results in any way other than for statistical reporting and analysis. **Initials:** _____
- I will use appropriate safeguards to prevent use or disclosure of the information other than as provided for by this agreement. **Initials:** _____
- I will ensure all members of the research team who have access to the NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Database through SEER*Stat have signed this agreement. **Initials:** _____

APPENDIX K – NPCR Research Data Use Agreement

- I will not attempt to link or permit others to link NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Data with individually identifiable records from any other dataset without CDC approval. **Initials:** _____

- I will not attempt to use the analytic results or permit others to use them to learn the identity of any person or establishment included in any dataset. **Initials:** _____

- I will take the following actions if the identity of any person or establishment is discovered inadvertently:
  - Make no use of this knowledge.
  - Notify CDC by sending an e-mail to uscsdata@cdc.gov.
  - As requested by CDC, safeguard or destroy the information that identifies an individual or establishment.
  - Inform no one else of the discovered identity. **Initials:** _____

- I will make every effort to release all statistical information in such a way as to avoid inadvertent disclosure by:

  - Ensuring that no data on an identifiable case can be derived through subtraction or other calculation from the combination of tables in the given publication. **Initials:** _____

  - Ensuring that no data permit disclosure when used in combination with other known data. **Initials:** _____

  - Not disclosing or otherwise making public data on any unit smaller than 16. If the total number of cases in a cell is fewer than 16, the cell data will be suppressed in oral and written presentations. **Initials:** _____

- I have read the data documentation file and have an understanding of the data available in the database and the restrictions related to their use. If I have questions regarding my analytic approach, I will contact CDC NPCR (uscsdata@cdc.gov) for assistance. **Initials:** _____

- I am familiar with the use of **SEER*Stat** in analyzing data or will complete the needed training. **Initials:** _____

- I understand that I am responsible for the results of my own analysis. The findings and conclusions resulting from the analysis of these data are those of the authors and do not necessarily represent the official position of CDC. **Initials:** _____

- I will acknowledge central cancer registries whenever data are presented, released, or published by including the following (or similar) statement:

  > These data were provided by central cancer registries participating in CDC's National Program of Cancer Registries (NPCR) and/or NCI's Surveillance, Epidemiology, and End Results (SEER) Program and submitted to CDC and NCI in November {date}. **Initials:** _____

- As appropriate, I will cite the data –

APPENDIX K – NPCR Research Data Use Agreement

For the {date}-{date}*database:* National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Database, Nov {year} submission ({year}-{year}), United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released {date}, based on November {year} submissions. Available at www.cdc.gov/cancer/public-use.

For the {year}-{year} *database:* National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: NPCR and SEER Incidence – U.S. Cancer Statistics Public Use Research Database, Nov {year} submission ({year}-{year}, United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released {date}, based on November {year}submissions. Available at www.cdc.gov/cancer/public-use.

Initials: _____

_____ _____

Signature **Date**

Name: _____

Title and organization: _____

SEER*Stat username: _____

Telephone number: _____ E-mail address: _____

Please complete the above fields, sign and date the agreement, and email both pages to uscsdata@imsweb.com.

APPENDIX L – NPCR Data at the NCHS RDC Q&A

Can you summarize what CDC is planning to do?
CDC uses the National Center for Health Statistics (NCHS) Research Data Center (RDC) as a mechanism for researchers outside of the Division of Cancer Prevention and Control (DCPC) to request and gain access to the Restricted-Access NPCR/SEER data for research purposes. The data will be available through the NCHS RDC only after the standard data quality reviews that occur as part of the preparation for USCS and State Cancer Profiles.

The use of the NCHS RDC to manage data access will provide the highest level of data security and protection of confidentiality that is available for analysis of data. Any researcher must submit a proposal which will be reviewed and approved by CDC and representatives from the participating central cancer registries (CCRs) before any data analysis begins. Trained data analysts at the NCHS RDC create a dataset that is customized to each analysis. The researcher can run his or her own statistical analysis or have the NCHS RDC analyst run the analysis. The NCHS RDC analyst reviews all output from statistical analysis to ensure that the researcher only conducts analyses relevant to the approved protocol and that small cell sizes are suppressed. Absolutely no individual level data will leave the NCHS RDC facilities.

What is National Center for Health Statistics (NCHS)?
NCHS is one of the national centers at CDC and is located in Hyattsville, Maryland. As the Nation's principal health statistics agency, staff at NCHS compile statistical information to guide actions and policies to improve the health of our people. More information about NCHS is available at: http://www.cdc.gov/nchs/about.htm.

What is the Research Data Center (RDC)?
The NCHS RDC began in 1998 and has a long-standing history of managing access to health and vital statistics data through a rigorous proposal review process as well as review of the statistical output. The NCHS RDC mission is to give public access to the full range of health and vital statistics data, while protecting the confidentiality of the respondents and institutions that collected the information. There have been no breeches of confidentiality for data access through the NCHS RDC.

The NCHS RDC houses sensitive, but not classified, data. It allows access to individual data without the possibility of disclosure of identifying information. The NCHS RDC offers statistical, programming, and consulting expertise to facilitate the data analysis for research.

The NCHS RDC is a data hosting center, not a data repository. The data extracts that are hosted on the NCHS RDC are tailored specifically to the proposal and have a research life cycle. Once the analysis is completed, the data extract is archived for 2 years and then destroyed.

There are currently three modes of access through the NCHS RDC, each with specific restrictions. Access is available on-site at two locations (Hyattsville, MD and Atlanta, GA), nine Census RDCs, or through remote electronic access. More information about the NCHS RDC is available at: http://www.cdc.gov/rdc/

APPENDIX L – NPCR Data at the NCHS RDC Q&A

Why does CDC use the NCHS RDC?

Maintaining confidentiality is the primary objective of the NCHS RDC. Staff at NCHS RDC have statistical expertise to address confidentiality and disclosure risk. Using the NCHS RDC will allow CDC to comply with the Assurance of Confidentiality [308(d)] that was obtained for the NPCR-CSS data. All researchers must take confidentiality orientation, complete confidentiality forms, and review the disclosure manual, all of which outline practices that are essential to protecting the data and preventing disclosure of confidential information. Additionally, data housed at the NCHS RDC are not subject to the Freedom of Information Act (FOIA). More information about confidentiality is available at: http://www.cdc.gov/rdc/B4ConfiDisc/CfD400.htm.

What is the research proposal process?

The NCHS RDC has a rigorous review process for analyses proposed by any researchers wanting to use RADS data. All proposals will be evaluated by a Review Committee consisting of: the NCHS RDC Director, the Confidentiality Officer, the assigned NCHS RDC analyst, and NPCR representatives. The iterative review and comment process may take 6 to 8 weeks.

Through this process, the NCHS RDC staff, the NPCR staff, and the CCR staff will fully understand the intended analysis and will be able to provide any needed direction or restrictions on the analysis and describe any limitations in what is proposed. It will be possible for CDC and participating registries to disapprove a proposal. However, guidance and re-direction as needed should be the norm. More information about the review process is available at: http://www.cdc.gov/rdc/B3Prosal/PP300.htm.

Once a proposal has been approved, the NCHS RDC offers a secure environment for data analyses and has processes in place to review data output for small cell sizes. This will ensure that the NPCR suppression rules are properly applied. Through the NCHS RDC, the user can conduct analyses and have remote access to data but cannot download the individual record level data or obtain counts for inappropriately small cell sizes.

The use of the NCHS RDC to host the NPCR data is a win-win opportunity because of the confidence in knowing that the data are being used correctly and safely, while at the same time making the data available for external researchers in an appropriate way. In addition, this approach will not overtax resources here in the Branch or in the CCRs. The NCHS RDC provides a level of data control beyond that of any other data access system used for registry data.

Who has access to the data and at what level?

The NCHS RDC analysts will have access to the individual record level data since it is easier to create an analytic dataset using these data. The NCHS RDC analysts will be bound by the same data use agreements that CDC staff sign on an annual basis. Researchers with approved proposals will be able to conduct analyses through the NCHS RDC on the created dataset or have the NCHS RDC analyst do the analysis for them. However, they will not be able to download

any part of the data from the NCHS RDC. Any additional variables that were not included in the original analysis proposal will need a separate approval process.

Note that this is different from the process that NPCR has used in the past where researchers with approved proposals would have direct access to the dataset itself including the ability to download the data and create a listing of individual record level data and all variables in the RADS.

Researchers have several possible modes of access to the data set created for their specific research proposal. More information is available at: http://www.cdc.gov/rdc/B2AccessMod/ACs200.htm.

When a researcher conducts an analysis, what type of output will he or she get?
If a researcher is on-site at the NCHS RDC, he or she can save the results on the hard drive of the NCHS RDC computer. The NCHS RDC analyst will review the output for disclosure then either load the output onto a flash drive supplied by the researcher or e-mail the output files to the researcher. If a researcher is accessing the NCHS RDC remotely, he or she will send program by e-mail and, after disclosure review by the NCHS RDC analyst, will receive the output files by e-mail. No individual record level data are released to the researcher.

Will the CCRs be able to decide whether their data will be available through the NCHS RDC?
Starting with DP17-1701, participation in all CDC-created and hosted analytic datasets and web-based data query systems, as outlined in the annual NPCR-CSS Data Release Policy, is a required strategy. [DP17-1701, 2. CDC Project Description, a. Approach, iii. Strategies and Activities, Program 3: National Program of Cancer Registries (NPCR) – Component 1, Strategy 3 Cancer Data and Surveillance (Domain 1), Data Submission (page 19)] Therefore, data from all CCRs meeting eligibility requirements are included. Data use is important to NPCR and for continued support of the registries.

Will the CCRs be able to decide if their county-identifying variable (County at Dx [NAACCR#90]) is to be available for use in the NCHS RDC?
Starting with DP17-1701, participation in all CDC-created and hosted analytic datasets and web-based data query systems, as outlined in the annual NPCR-CSS Data Release Policy, is a required strategy. [DP17-1701, 2. CDC Project Description, a. Approach, iii. Strategies and Activities, Program 3: National Program of Cancer Registries (NPCR) – Component 1, Strategy 3 Cancer Data and Surveillance (Domain 1), Data Submission (page 19)] Therefore, data from all CCRs meeting eligibility requirements are included. County data will be used only in approved analyses and in the following ways:

- Used as a linkage variable (linkage to census data, for example) only by the NCHS RDC analyst. The county variable will not be available to the researcher but the NCHS RDC analyst would use it to create a linked dataset and then remove the county variable.

- Included as a confounder or other control variable, but no data are presented by county. The NCHS RDC analyst will create dummy variables to mask the actual county name.

- 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). It will be possible for the NCHS RDC analyst to create these areas for the researcher.

Previous data release policies indicate that the project proposals for RADS would be reviewed by the RADS working group, facilitated by CDC with representation by the CCRs. Does this procedure change now that the NCHS RDC is used?

The CCRs will still have input on the RADS proposals. The NCHS RDC review process also includes the NCHS RDC analyst and the confidentiality officer, who will be responsible mainly for disclosure review to ensure that we abide by the 308(d) assurance of confidentiality obtained for NPCR-CSS. More information about the NCHS RDC review process is available at: http://www.cdc.gov/rdc/B3Prosal/PP340.htm.

NPCR will obtain comments on each proposal from CCRs through the NPCR Central Cancer Registry Council.

Will SEER data be included for analysis or will the data be limited to NPCR data?
Yes. Both NPCR and SEER data may be accessed through the NCHS RDC.

Will the NCHS RDC staff have access to SEER*Prep and SEER*Stat?
Yes. NPCR staff have worked with NCHS RDC staff to provide appropriate training for these data preparation and analysis tools. All previous analyses performed on the data at the NCHS RDC have required a SAS dataset and is the primary data source being used.

Will researchers have access to SEER*Stat?
Yes. It is expected that researchers will know the basics of the analyses that they wish to carry out. NCHS RDC staff will be available for limited consultation. Since cell phones or access to the Internet are not available inside the NCHS RDC, all SEER*Stat tutorials (http://seer.cancer.gov/seerstat/tutorials/) would need to be completed beforehand.

What suppression rules will be used for the RADS?
The same suppression rules that are used for *United States Cancer Statistics.* More detailed information is available at:
https://www.cdc.gov/cancer/npcr/uscs/technical_notes/stat_methods/suppression.htm.

In addition, the suppression rules for Asians/Pacific Islanders (A/PI) and American Indians/Alaska Natives (AI/AN) will also apply. The data for A/PI and AI/AN will be presented only for states or counties with at least 50,000 population because of concerns regarding possible misclassification of race data and the relatively small sizes of these populations in the United States.

Wouldn't it be better for researchers to contact CCRs directly for linkage studies? CDC doesn't collect personal identifiers like name or social security number.

Yes, it would be best for researchers to contact CCRs directly for linkage studies that require individual identifiers. However, valuable public health research can be conducted with access to county-level data. Examples include linkage with U.S. Census data for socioeconomic analyses, or to examine regional differences in the prevalence of a specific cancer

Will IRB review be required for each proposal? If not, will NCHS require the researcher to obtain IRB approval before they submit their proposal?
The NCHS RDC has an umbrella ethics review board (ERB) protocol that covers CDC employees and can be extended to external researchers. The principal investigator and all research team members who come in contact with the data must take the confidentiality orientation and complete the confidentiality forms. One of the confidentiality forms is the designated agent form (http://www.cdc.gov/rdc/Data/B4/DesignatedAgent.pdf), which extends the ERB to cover external researchers.

Note that the ERB protocol serves the same function as an institutional review board (IRB) protocol. At CDC, there is one office that coordinates the submission and tracking of human research protocols. However, other centers such as NCHS and the National Institute of Occupational Safety and Health, have different names for these review boards: Research Ethics Review Board (ERB) at NCHS and Human Subjects Review Board (HSRB) at NIOSH.

Researchers may choose to obtain an IRB from their own institution, but it will not be a requirement in the application process given the ERB extension that the NCHS RDC provides.

Does access to the RADS cost anything?
No. CDC covers the cost of analyzing RADS through the NCHS RDC.

As more researchers become aware of the RADS, they may want access to additional variables that CCRs submit to CDC. How will this process be handled?
The addition of new variables in RADS will be discussed with CCRs prior to their inclusion in the data release policy, which is updated annually.

How is access to the comparative effectiveness research (CER) dataset managed?

APPENDIX L – NPCR Data at the NCHS RDC Q&A

Access to the CER dataset are managed through the same NCHS RDC process. The proposal process will not differ except that staff from the Specialized Registries funded for CER data collection will review these proposals.

APPENDIX M – Data Items for Restricted-Access Dataset (RDC)

The restricted access dataset are individual case-specific data derived from the NPCR-CSS dataset. The data is available to researchers at NCHS Research Data Centers as a SAS file. SAS files are created specifically for each project's needs. The data items that may be requested by researcher are listed below.

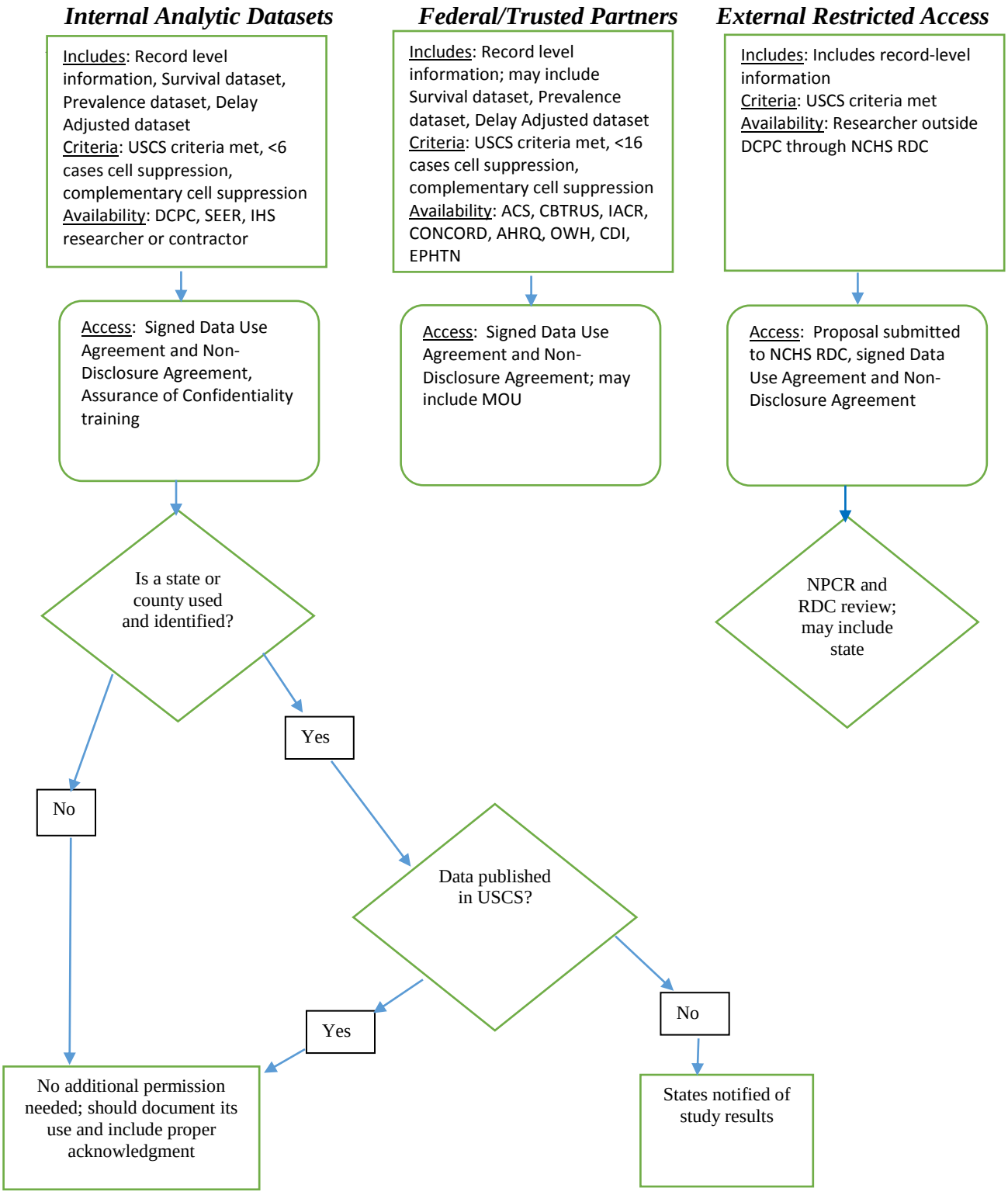
Variable Name
Alternate Patient ID Number
Address at Diagnosis – State
Address at Diagnosis – County*
USCS Standard
USCS9915
USCS1115
USCS9815
USCS0615
Address at Diagnosis – Census Region
Race 1
Race 2
Race Recode
Econ Status
State race eth suppress
Spanish/Hispanic Origin
NHIA Derived Hispanic Origin
IHS Link
Sex
Age at Diagnosis**
Age Recode
Birth Date***
Sequence Number – Central
Date of Diagnosis****
Primary Site
Laterality
Grade
Diagnostic Confirmation
Type of Reporting Source
Histologic Type ICD-O-3
Behavior Code ICD-O-3
Behavior Recode for Analysis
Primary Site Recode
Primary Site Recode with Mesothelioma and Kaposi Sarcoma
SEER International Classification of Childhood Cancer (ICCC) Recode
SEER Summary Stage 2000

APPENDIX M – Data Items for Restricted-Access Dataset (RDC)

SEER Summary Stage 1977
RX Summ – Surg Prim Site
Merged radiation
CS Extension
CS Lymph Nodes
CS Mets at DX
CS Site-Specific Factor 1
CS Site-Specific Factor 2
CS Site-Specific Factor 3
CS Site-Specific Factor 15
CS Site-Specific Factor 25
CS Version Input Original
CS Version Derived
CS Version Input Current
Derived SS2000
Merged Summary Stage 2000
Over-ride Age/Site/Morph
Over-ride SeqNo/DxConf
Over-ride Site/Lat/Sequence Number
Over-ride Site/Type
Over-ride Histology
Over-ride Report Source
Over-ride Ill-define Site
Over-ride Leuk, Lymphoma
Over-ride Site/Behavior
Over-ride Site/Lat/Morph
Alcohol-related cancers
HPV-related cancers
Obesity-related cancers
Physical activity-related cancers
Tobacco-related cancers

* 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)
**Age over 99 is recoded
***Only year is provided; if age is over 99, year of birth is recoded
****Day of diagnosis is not provided

APPENDIX O – Data Items for NPCR/SEER USCS Delay Adjusted Database



APPENDIX O – Data Items for NPCR/SEER USCS Delay Adjusted Database

NPCR-CSS Levels of Data Access
Public Use Datasets

CDC WONDER

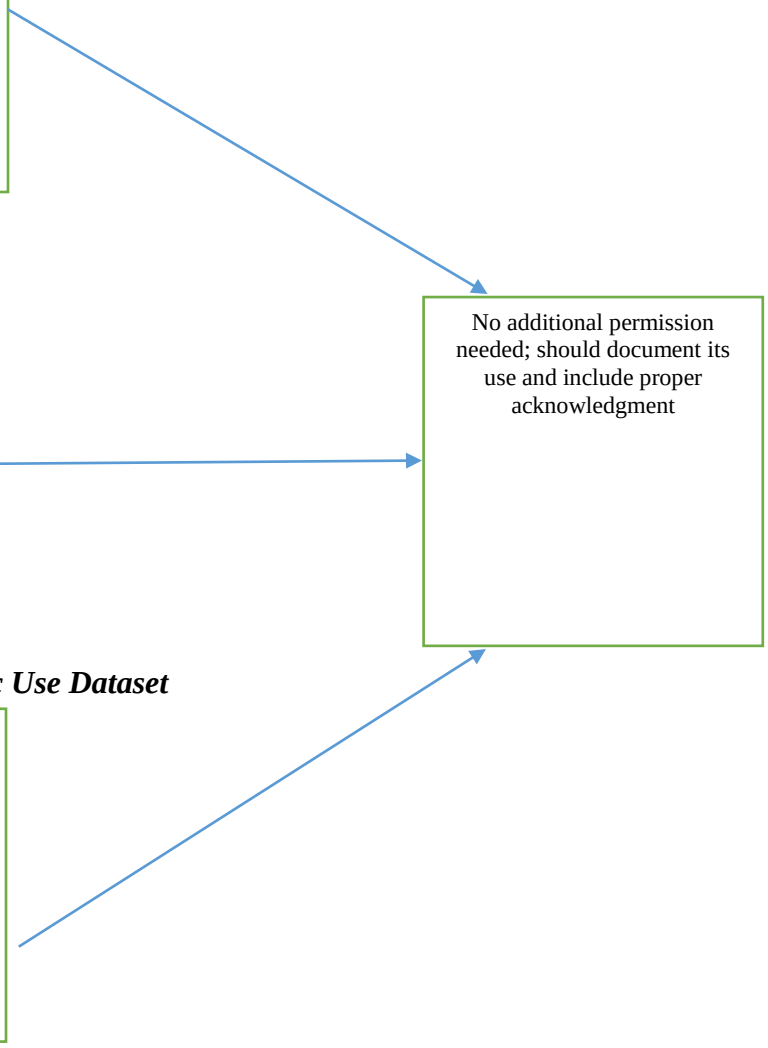
Includes: State and MSA levels, no record-level information
Criteria: USCS criteria met, permission provided on Dataset Participation Agreement, <16 cases cell suppression enforced
Availability: Public

State Cancer Profiles

Includes: State and county levels, no record-level information
Criteria: USCS criteria met, permission provided on Dataset Participation Agreement, <16 cases cell suppression enforced
Availability: Public

NPCR/SEER USCS Public Use Dataset

Includes: State record-level information, no case listing
Criteria: USCS criteria met, permission provided on Dataset Participation Agreement, <16 cases cell suppression enforced
Availability: Public after signed Data Use Agreement and Non-Disclosure Agreement, annual agreements required



APPENDIX O – Data Items for NPCR/SEER USCS Delay Adjusted Database

SEER*Stat Category	SEER*Stat Variable Name
Age at Diagnosis	Delay age Age recode with single ages and 85+
	Age recode with <1 year olds
Race, Sex, Year Dx, Registry, County	Sex
	Year of diagnosis
	Addr at DX – state
	County at DX
	State-county
Required Delay Fields	Delay factorNPCR project flag
	Delay siteEconomic status 2015
	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)
	Delay race (All, Race recode (White, Black, AIAN, CHSDA, API, Hisp, Non-HispOther)
	Year of diagnosis
Site and Morphology	Primary Site – labeled
	Histologic Type ICD-O-3
	Behavior Code ICD-O-3
	Grade
	Diagnostic confirmation
	ICD-O-3-Hist/behavior, labeled
	ICD-O-3-Hist/behavior, malig, labeled
	Site recode ICD-O-3/WHO 2008
	ICCC site recode ICD-O-3/WHO 2008
Site and Morphology	Behavior recode for analysis derived/WHO2008
Stage – LRD [Summary and Historic]	Derived SS2000
	SEER Summary Stage 2000
	Merged Summary Stage 2000
Extent of Disease – CS	CS Site-Specific Factor 1
	CS Site-Specific Factor 2
	CS Site-Specific Factor 15
	Laterality
Cause of Death (COD) and Follow-up	Survival months – presumed alive
	Survival months flag – presumed alive
	Cause of death (ICD-10)
	ICD revision number
	Vital status
	Follow-up source central
	COD exclusion flag

APPENDIX O – Data Items for NPCR/SEER USCS Delay Adjusted Database

SEER*Stat Category	SEER*Stat Variable Name
	Original vital status
	Vital status recode (study cutoff used)
	Cause of death recode
	COD recode with Kaposi and mesothelioma
Multiple Primary Fields	Sequence number - central
Race and Age (case data only)	Age at Diagnosis
	Race 1
	NHIA derived Hispanic origin
Dates	Presumed alive year of last contact recode
	Presumed alive month of last contact recode
	Presumed alive day of last contact recode
	Year of birth
	Month of diagnosis
	Day of diagnosis
	Original day of last contact
	Original month of last contact
	Original year of last contact
	Original year of diagnosis
	Original day of diagnosis
	Original month of diagnosis
Other	Type of Reporting Source
User-Specified	EDPMDE LinkVar

APPENDIX P – Data Items for NPCR Prevalence Database

SEER*Stat Category	SEER*Stat Variable Name
Age at Prevalence Date	Age at Prevalence Data (Calculated)
Age at Diagnosis	Age recode with single ages and 85+
Race, Sex, Year Dx, Registry, County	Sex
	Year of diagnosis
	Addr at DX – state
	County at DX
	State-county
	NPCR project flag
	Economic status 2015
	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)
	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)
	State
	County
	Race recode (White, Black, Other)
	Year of diagnosis
Site and Morphology	Primary Site – labeled
	Histologic Type ICD-O-3
	Behavior Code ICD-O-3
	Grade
	Diagnostic confirmation
	ICD-O-3-Hist/behavior, labeled
	ICD-O-3-Hist/behavior, malig, labeled
	Site recode ICD-O-3/WHO 2008
	ICCC site recode ICD-O-3/WHO 2008
	Behavior recode for analysis derived/WHO2008
Stage – LRD [Summary and Historic]	Derived SS2000
	SEER Summary Stage 2000
	Merged Summary Stage 2000
Extent of Disease – CS	CS Site-Specific Factor 1
	CS Site-Specific Factor 2
	CS Site-Specific Factor 15
	Laterality
Cause of Death (COD) and Follow-up	Survival months – presumed alive
	Survival months flag – presumed alive
	Cause of death (ICD-10)
	ICD revision number
	Vital status

APPENDIX P – Data Items for NPCR Prevalence Database

SEER*Stat Category	SEER*Stat Variable Name
	Follow-up source central
	COD exclusion flag
	Original vital status
	Vital status recode (study cutoff used)
	Cause of death recode
	COD recode with Kaposi and mesothelioma
Multiple Primary Fields	Sequence number - central
Race and Age (case data only)	Age at Diagnosis
	Race 1
	NHIA derived Hispanic origin
Dates	Presumed alive year of last contact recode
	Presumed alive month of last contact recode
	Presumed alive day of last contact recode
	Year of birth
	Month of diagnosis
	Day of diagnosis
	Original day of last contact
	Original month of last contact
	Original year of last contact
	Original year of diagnosis
	Original day of diagnosis
	Original month of diagnosis
Other	Type of Reporting Source
User-Specified	EDPMDE LinkVar